Reinforcement learning in accelerators

Are we there yet?

Simon Hirlaender
Team lead: Smart Analytics und Reinforcement Learning
IDA Lab
Artificial intelligence and Human Interfaces
Digital and Analytical Sciences
University of Salzburg

Outline

- Motivation for RL and intro to RL
- What is CERN and why RL is interesting there
- History of RL and examples
- Conclusion and open questions

Outline

- **Motivation for RL and intro to RL**
- What is CERN and why RL is interesting there
- History of RL and examples
- Conclusion and open questions

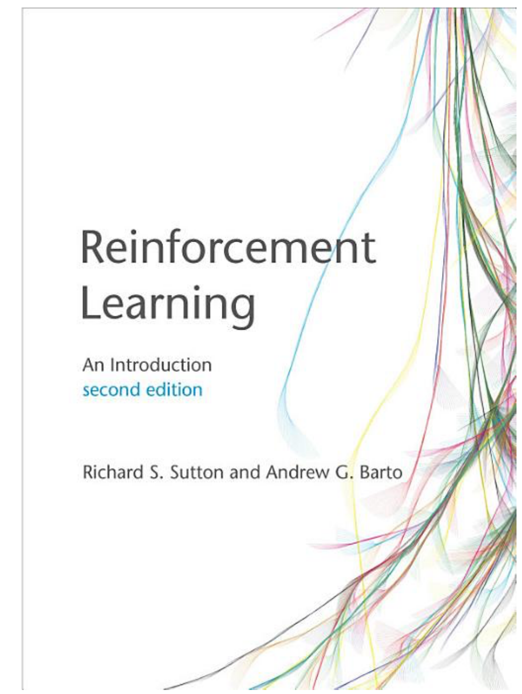
Recently I read in the NY times…

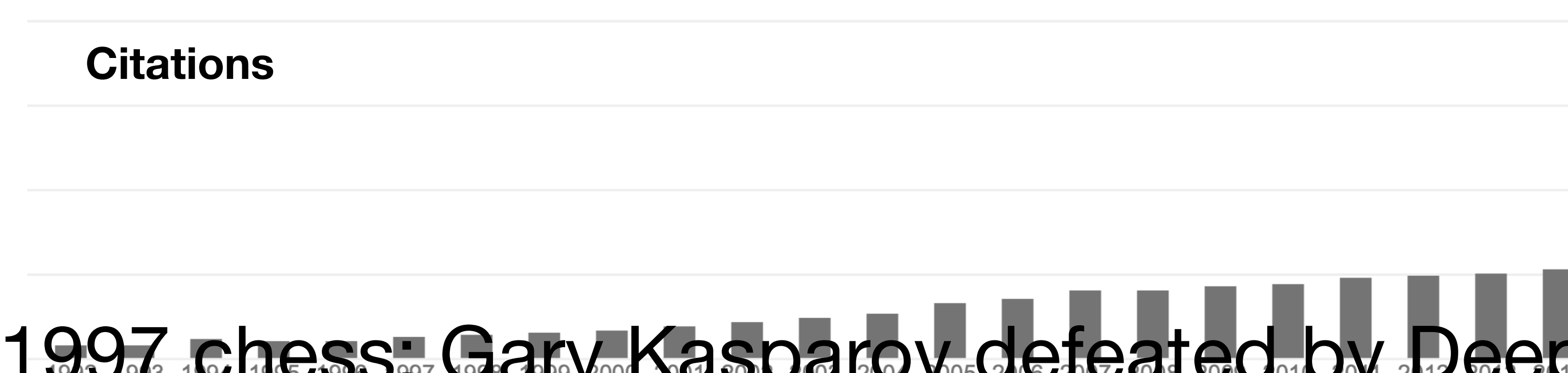
- *The Navy revealed the embryo of an electronic computer that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.*
- From 1958 referring to the perceptron by Rosenblatt
- Let to a boost of AI

https://www.nytimes.com/1958/07/08/archives/new-navy-device-learns-by-doing-psychologist-shows-embryo-of.html

2016: a milestone in artificial intelligence

Go: Lee Sedol was defeated by AlphaGo - using reinforcement learning





1997 chess: Gary Kasparov defeated by Deep Blue - (rule based)

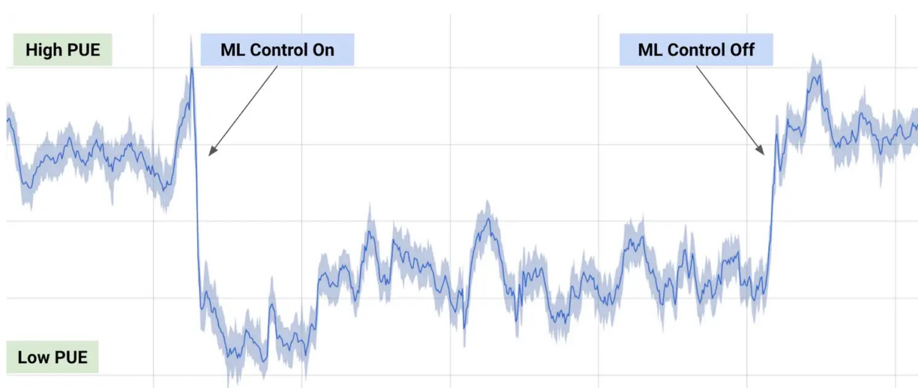
2018 @ Openai: solving Rubik's Cube with a Robot Hand

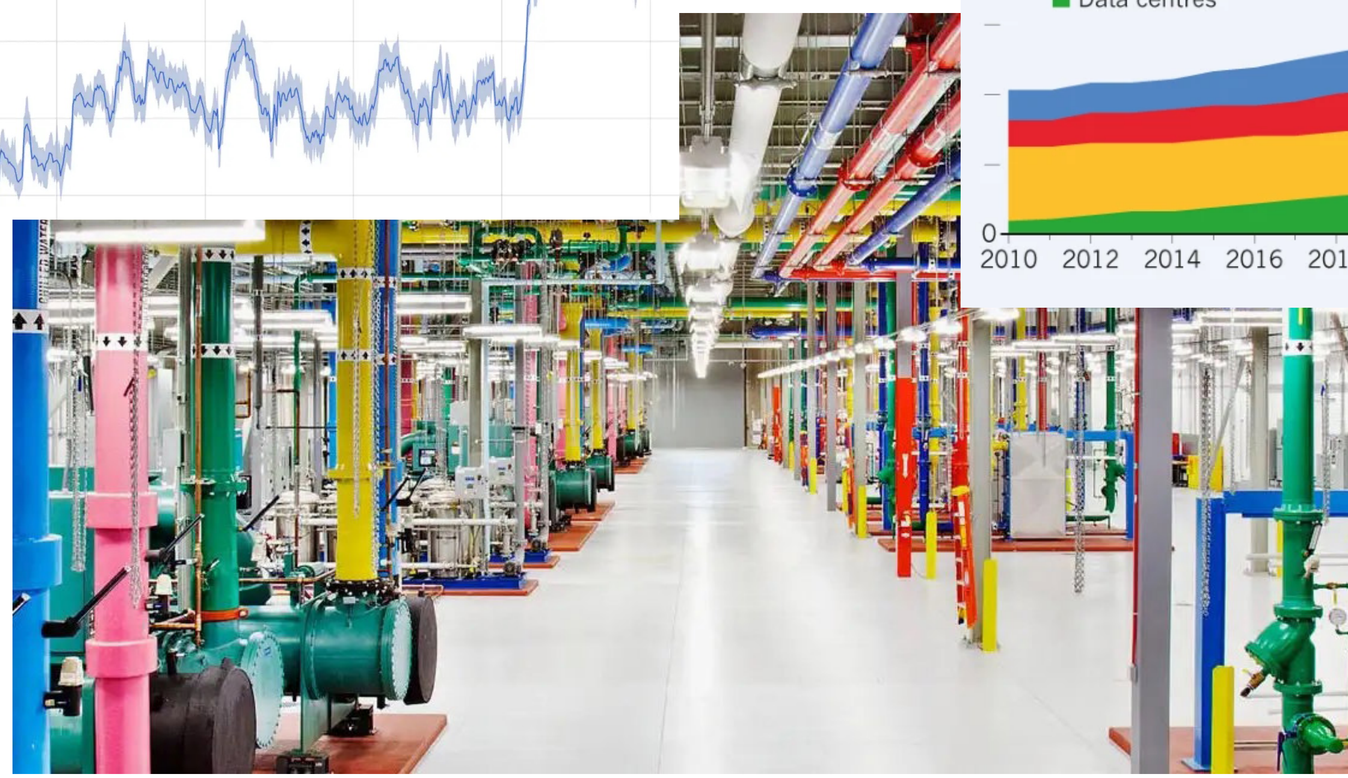
- RL goes beyond what we can engineer by hand

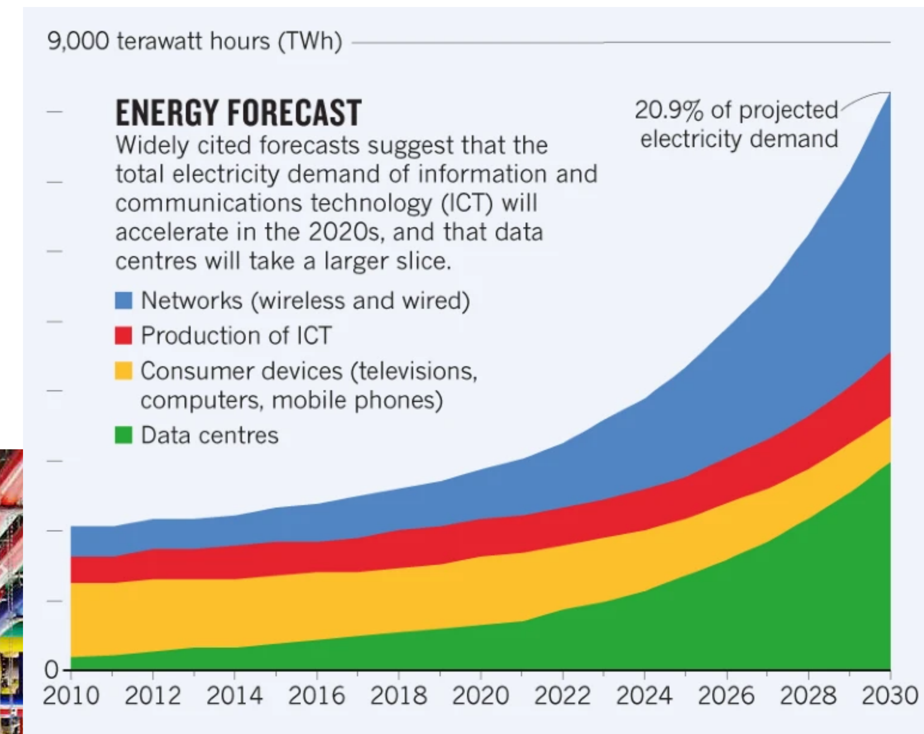
https://www.youtube.com/watch?v=x4O8pojMF0w

2018 @ Google: reducing energy consumption

DeepMind AI Reduces Google Data Centre Cooling Bill by 40% - using RL





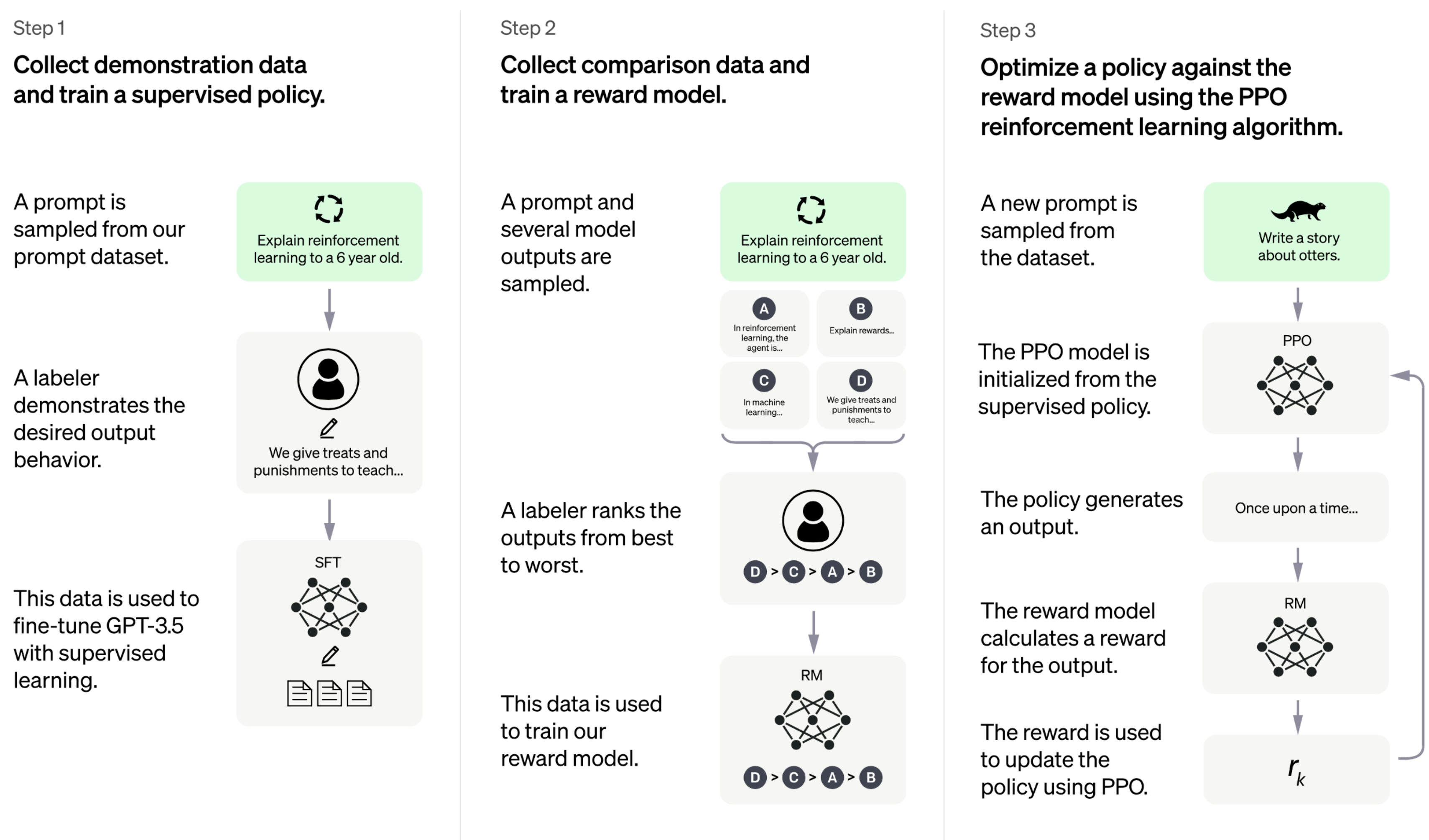


https://www.nature.com/articles/d41586-018-06610-y

2020: RL in industry (robotics)

https://covariant.ai/news/automation-upgraded-robotic-goods-to-person-picking

Now@Openai: Chat GPT (3.5)



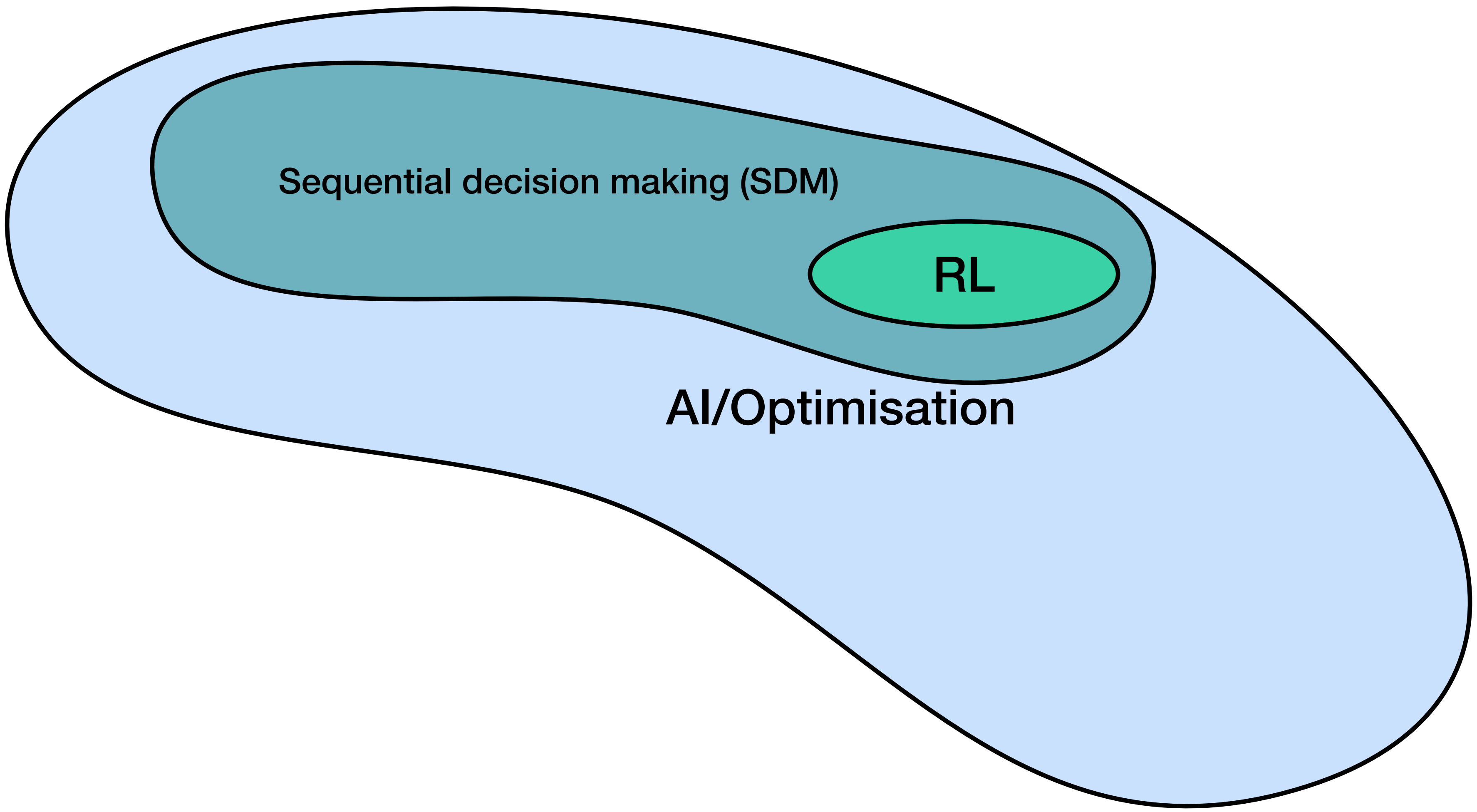
Huge societal impact ongoing

What is RL?

Addresses fundamental challenge of (artificial) intelligence and machine learning:

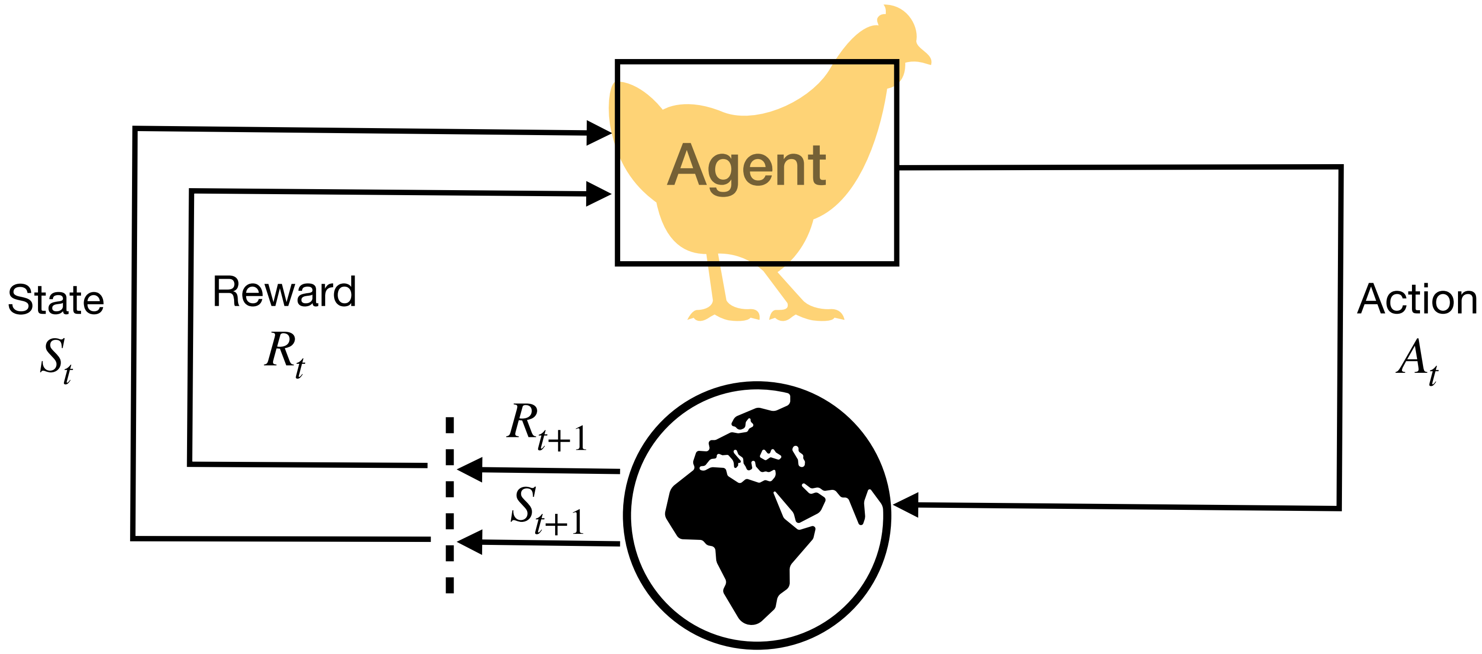
Learn how to make good decisions under uncertainty

Where does RL belong to?



How does RL work?

https://www.youtube.com/watch?v=spfpBrBjntg



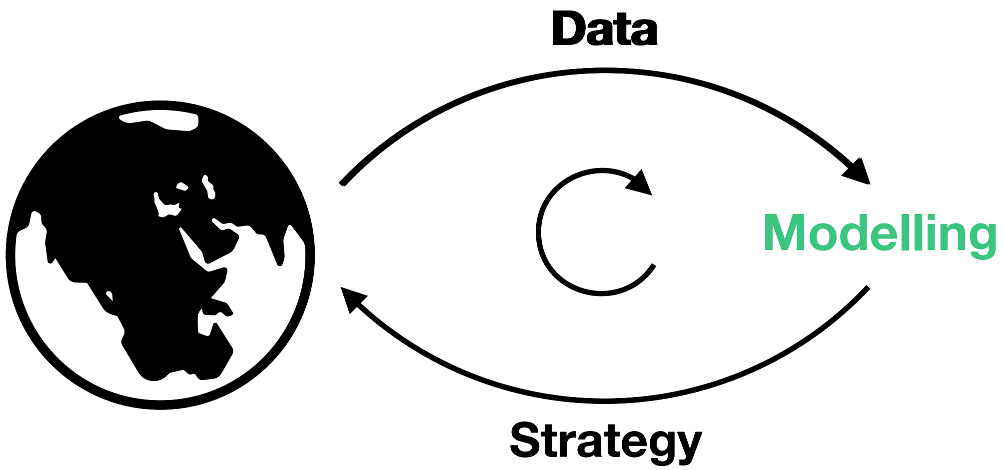
Learns from experience.
Goal: Maximising expected cumulative reward

$$\max \mathbb{E}[\sum_t R_t]$$

We try to find a function which tells us what a good decision is in every state s: $\pi(s) = a$

RL and decision theory

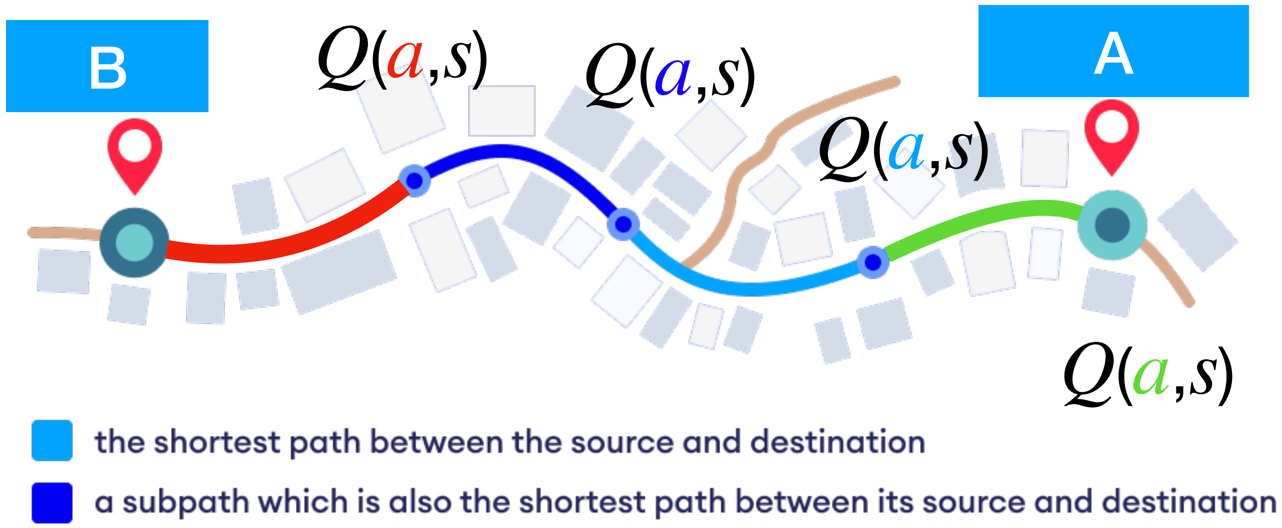
Information → decision → Information → decision → Information → …



- One step horizon offline RL ⇒ Prediction $\mathbb{P}(Y_i \mid X_i)$ - pattern recognition or supervised learning (SL)
- One step horizon RL ⇒ active Learning - e.g. system identification
- RL is a multi step **optimization** problem!

Bellman ~1957: dynamic programming

$$Q(a,s) = \mathbb{E}_{\pi}[\sum_t R_t \mid A_t = a, S_t = s]$$

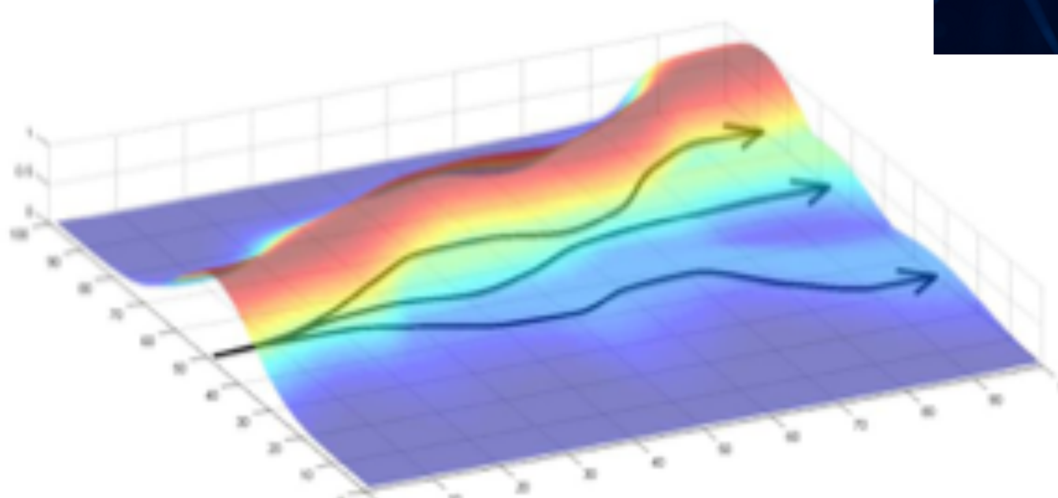


- Bellman idea:
  - ➡ Exact backwards recursion (if all transition probabilities are perfectly known) → unique solution for optimal policy
  - ➡ Stochastic approximation: central and novel to reinforcement learning - temporal-difference learning - using bootstrapping
  - ➡ Watkins 1989: Solving the control problem on small problems Q-learning
  - ➡ Basis of all **value-based** methods in RL - estimating the future reward of each state and construct a policy from there

Direct optimization of $\pi(a)$

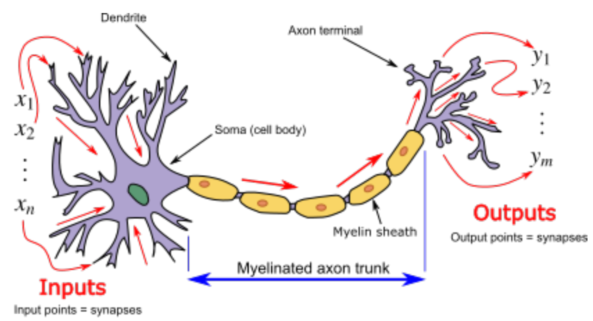
- Policy-based
- Derivative free optimization
- Random sampling
- Estimating the derivative

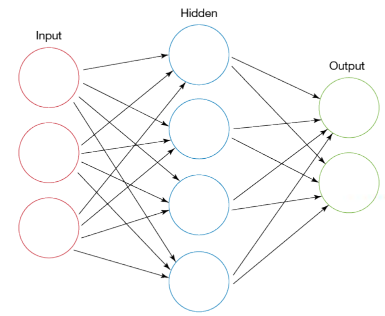
https://miro.medium.com/max/2000/1*ff14zY0i4mi3HPa6pCeF4g.png

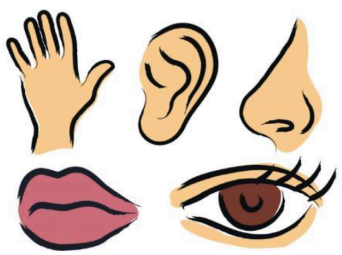


Adapted from Sergey Levine

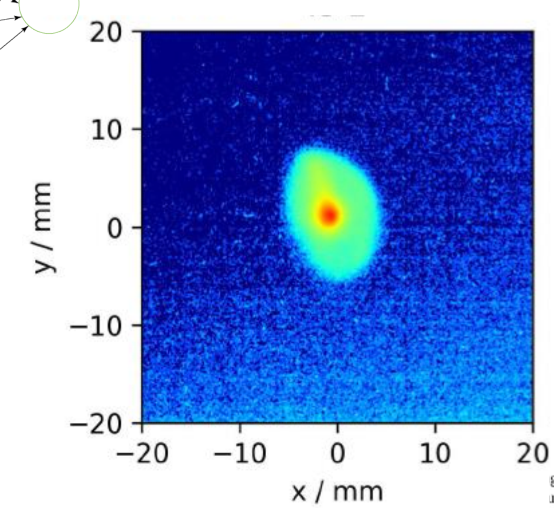
Why Deep Learning?



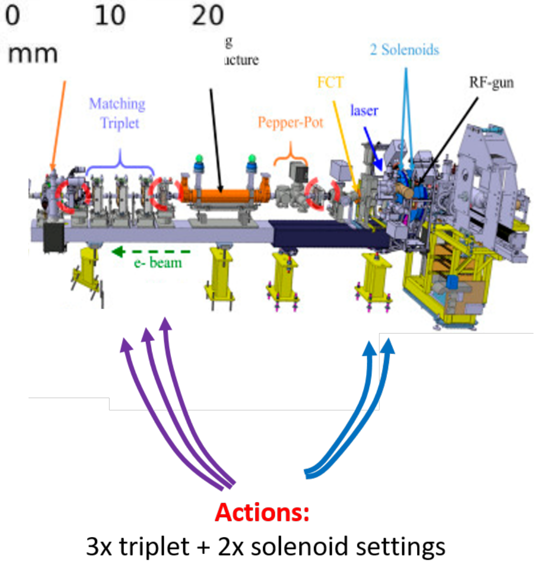


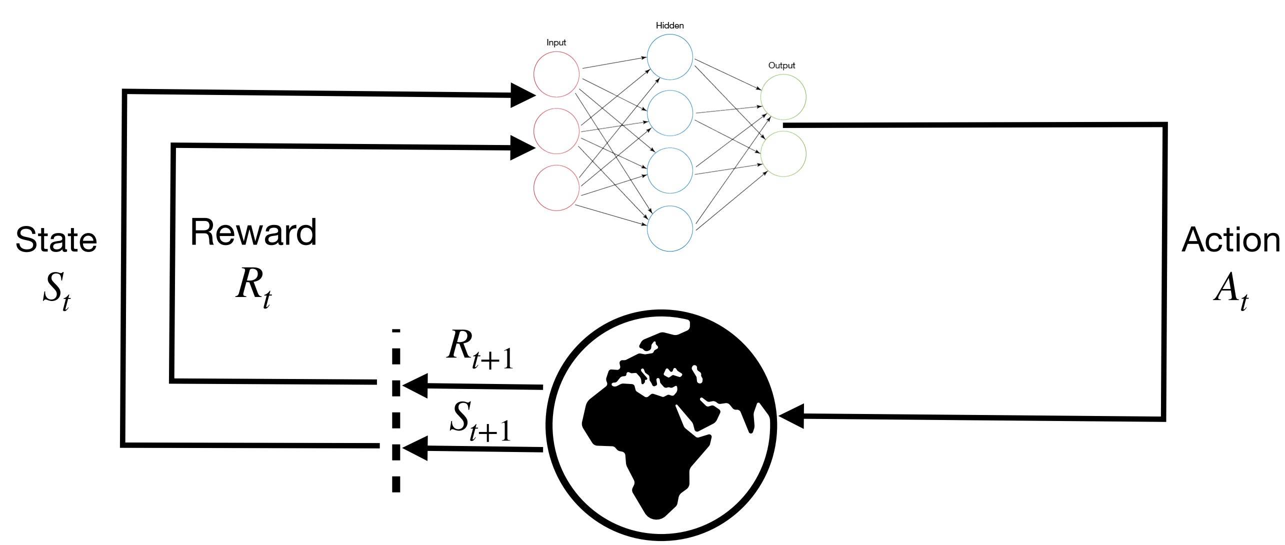


- Complex sensorial input



- Algorithms can select complex actions!





Modern Deep Reinforcement Learning

Make more stable!

Reduce variance!

Value based relies on Bellman updates

Policy based - stochastic optimisation

Actor/Critic

- Biased
- Can be offline
- Brittle
- No guarantees
- Deterministic
- Fast
- DDPG, TD3, DQN …

- High Variance
- Generally online
- Performance guarantees
- Local optima
- Stable
- Allows for stochastic policies
- AC, TRPO, PPO

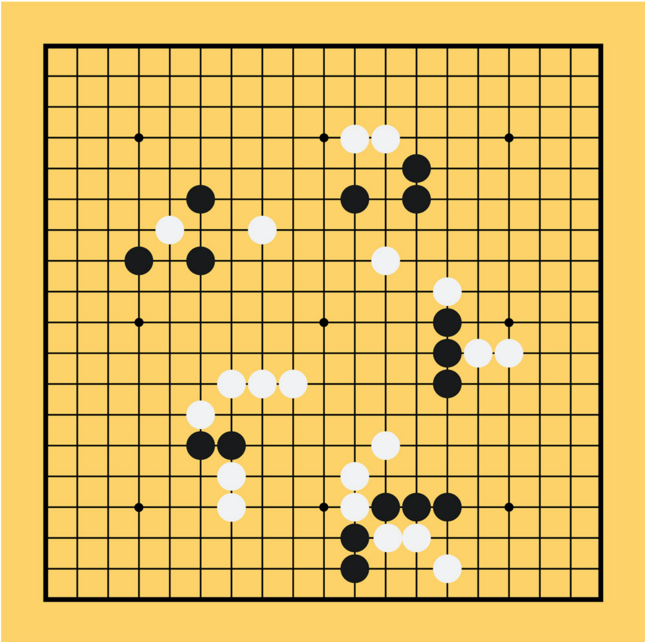
- SAC - Soft Q learning - stochastic
- D4PG - Distributional Q-Learning
- MPO

RL main points

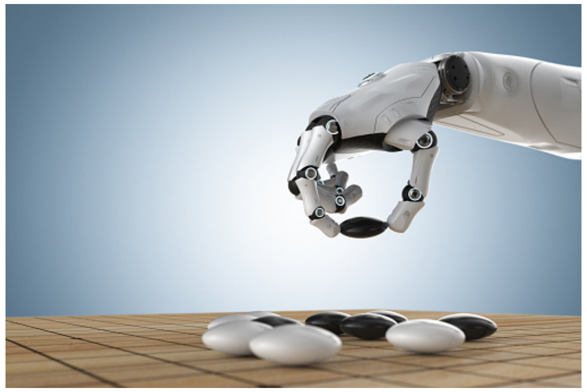
- **Learn a policy** $\pi(s) \mapsto a$ to maximise the expected return of a given problem **through experience**
- The **reward** (a scalar) - **designed by us** - tells the algorithm (the agent) - **what is good and what not**
- We have to **capture the problem well enough** so that a good policy can be learned
- RL can handle **delayed consequences**

Back to Go

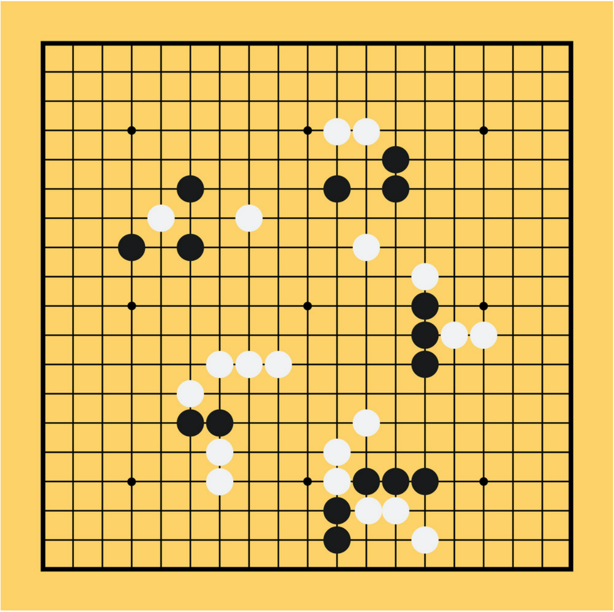
- AlphaGo Zero: 3,000 years of human knowledge in 40 days
- AlphaGo Zero played 4,9 million games against itself!
- **Only possible in simulations!**
- **Several hundred years of real play- apart from other problems**



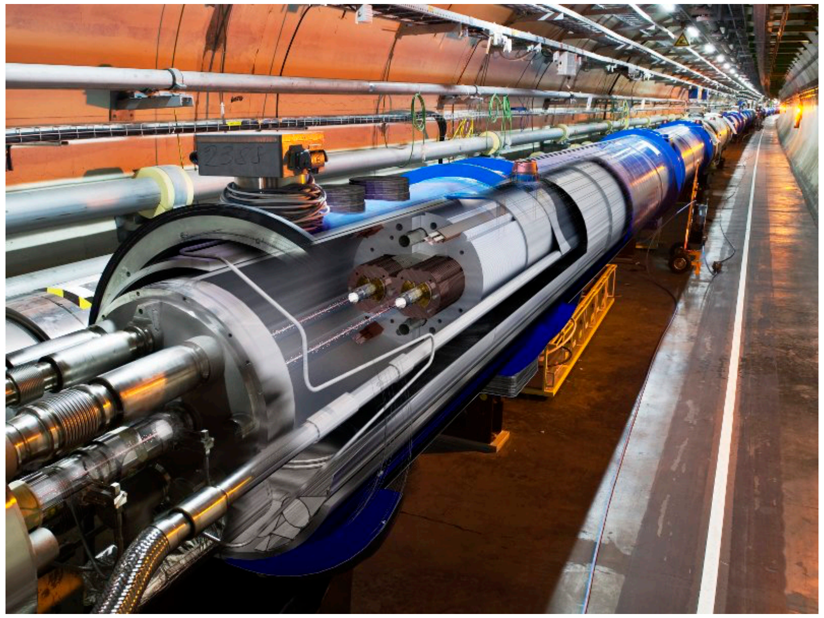
Real systems: as little data a possible



How to close the gap?

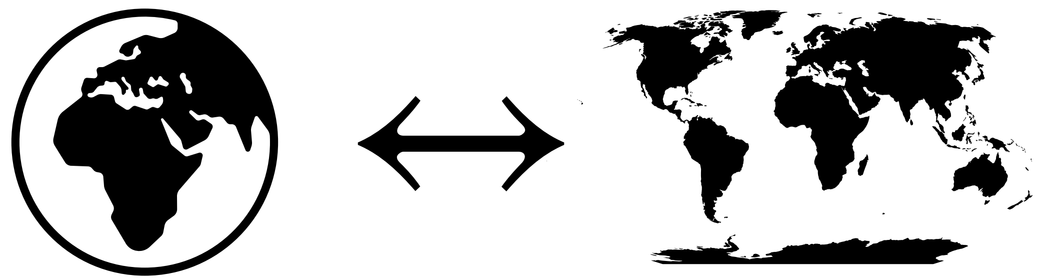


$$\overset{?}{\Rightarrow}$$



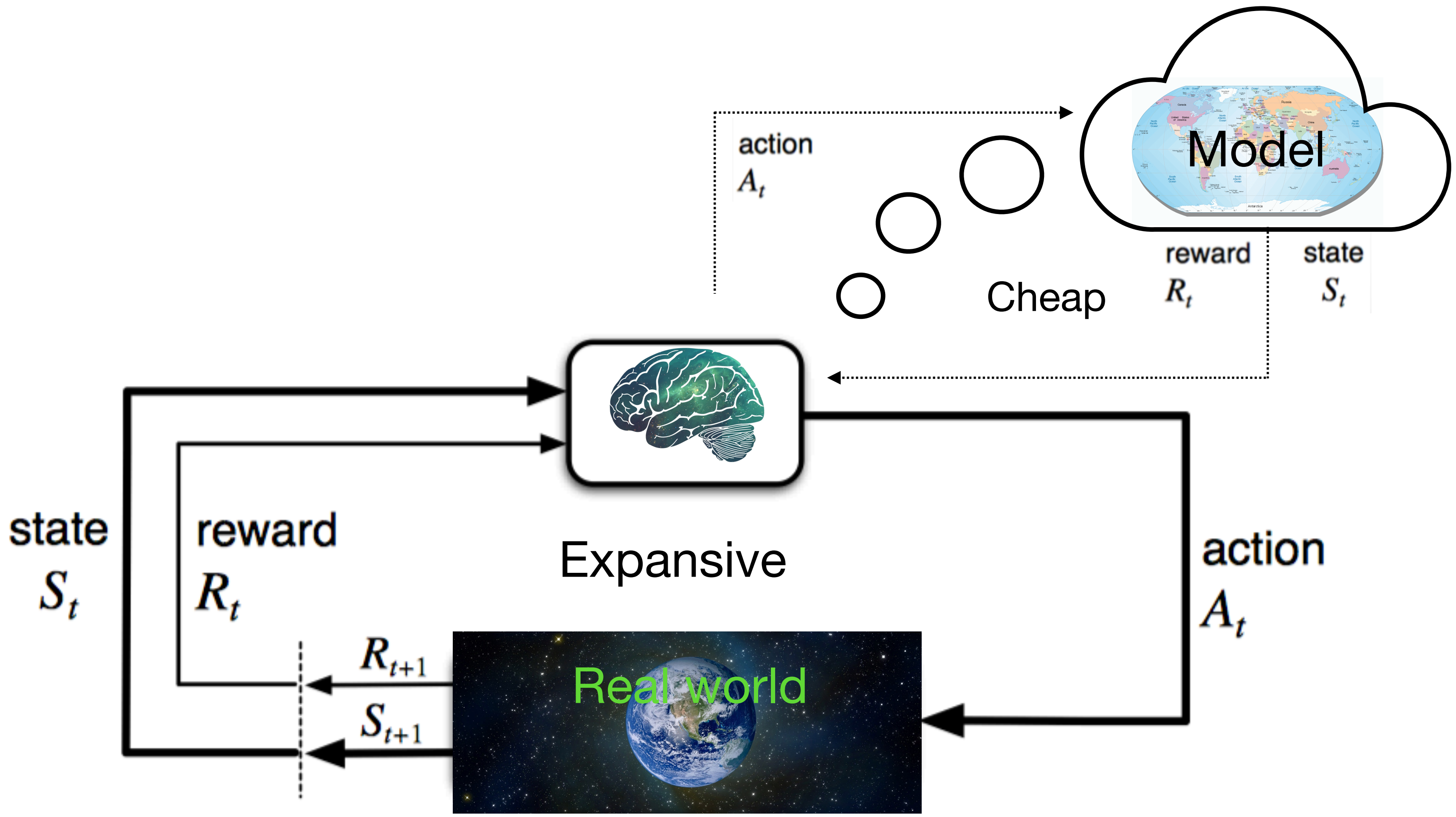
https://www.siliconrepublic.com/wp-content/uploads/2014/12/201411/large-hadron-collider.jpg

Why not just using a simulator?



- Approximate Markov decision process (MDP) via simulation
  - ➡ Can be complicated on its own
  - ➡ Accurate simulations are generally too slow or intractable at all
  - ➡ Imperfect model of MDP: transfer usually hard, long re-training
- Possible solutions: Replan, **learn a model** (then plan), do both…or novel paradigms as meta reinforcement learning

Model based RL - separation heuristic



Information → (Plan) → Decision → Information → (Plan) → Decision → …

Algorithmic challenges of RL in the real world

- Sample efficiency
- Stability/Guarantees
- Run time
- Hyperparameter tuning
- Exploration/Safety
- …
- Consequently, applying RL rather complicated
- Solutions are specific

Sample efficiency: how bad is it?

Generating data in real systems is generally limited

Planning

~ 100 interactions

~ 10e4 interactions

Optimisation

> 1e6 interactions

Model-based

Off-policy
Q-learning

Actor-critic

On-policy
Policy gradient

Evolutionary/
gradient free

Seconds

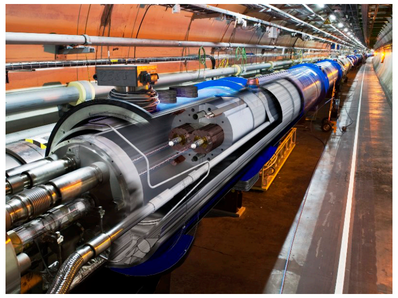
Minutes

One Hour

Hours

Tens of hours

Real time



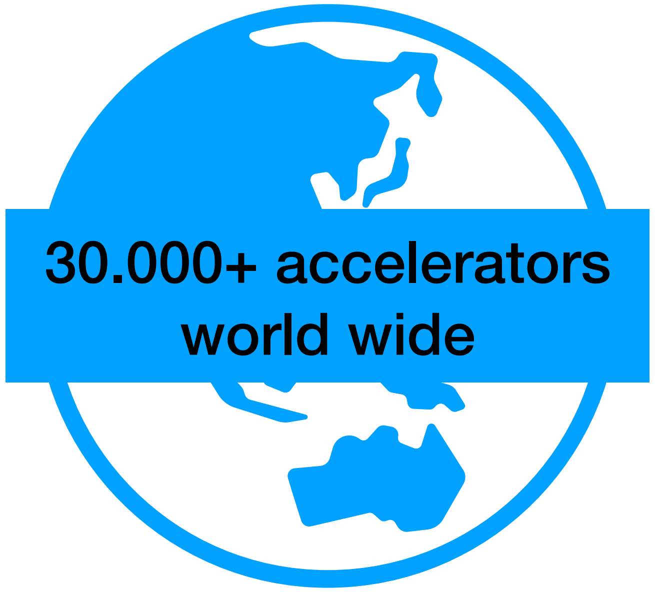
Outline

- Motivation for RL and intro to RL
- **What is CERN and why RL is interesting there**
- History of RL and examples
- Conclusion and open questions

The world of particle accelerators

- Machines generate charged energetic particle beams - many applications
- Complex set-up: many parameters to configure
- Optimisation algorithms and RL approaches are highly beneficial

Fundamental research (< 1 %)
- Fundamental physics
- Material studies
- Biology, chemistry



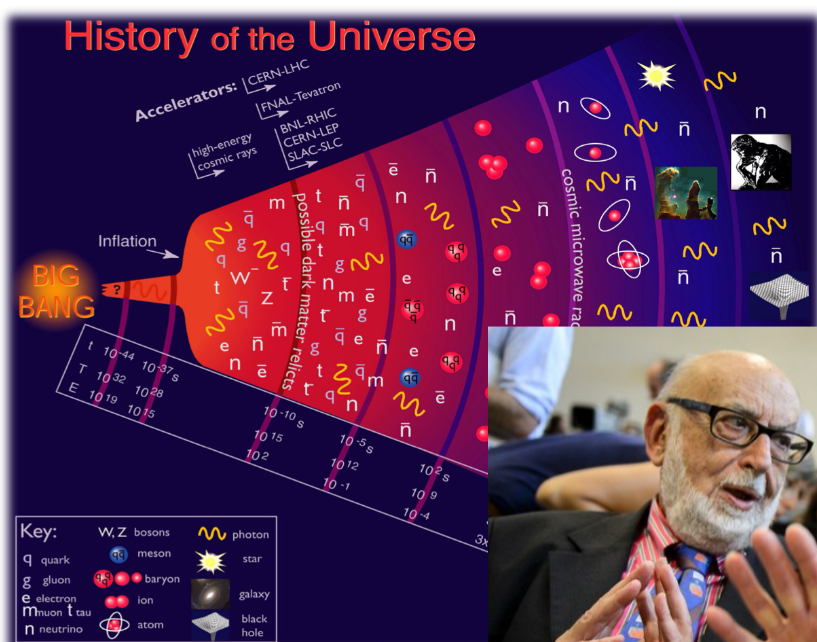
Security
- Cargo inspection
- Material characterisation

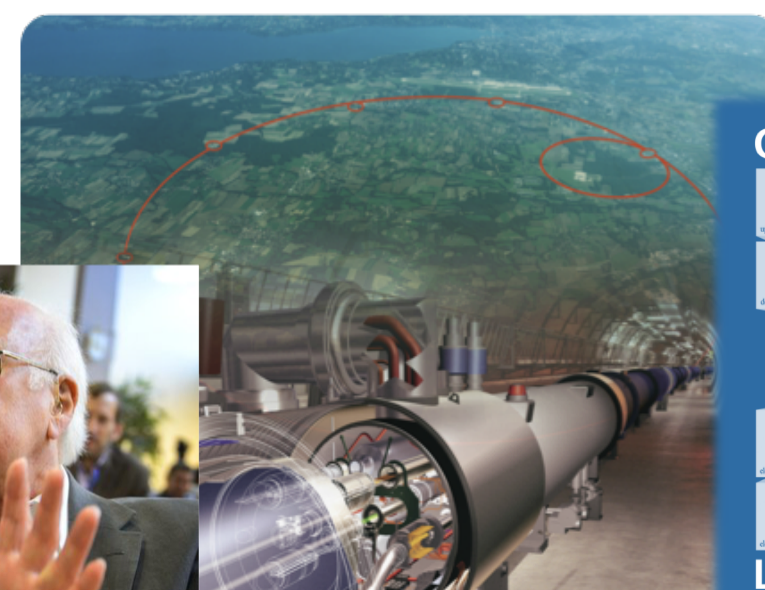
Industry
- Material / Surface/treatment
- E.g. computer chip production
- Sterilisation of food

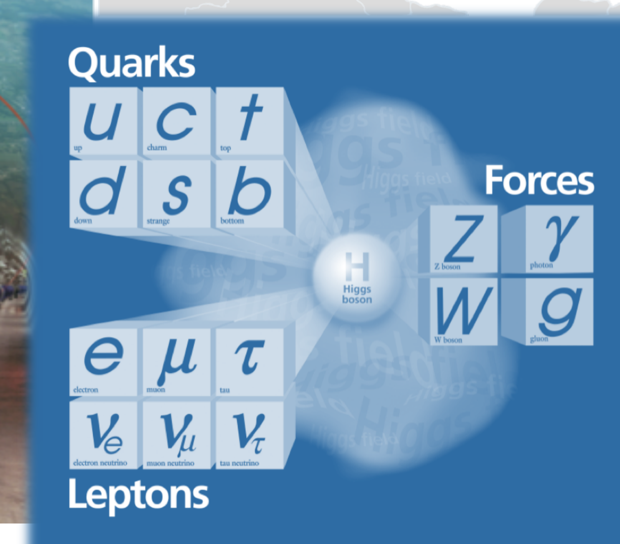
Medicine
- Isotop-production
- Cancer diagnosis and treatment industry

What is CERN?

- European Organization for Nuclear Research, founded in 1954, located near Geneva, Switzerland
- “Science for Peace”
- Largest particle physics lab in the world (12k+ users from 70+ countries)
- Mission: providing and operating particle accelerators and infrastructure for fundamental research in high-energy physics
- Current flagship: Large Hadron Collider (LHC), but there are many more accelerators and experiments at CERN





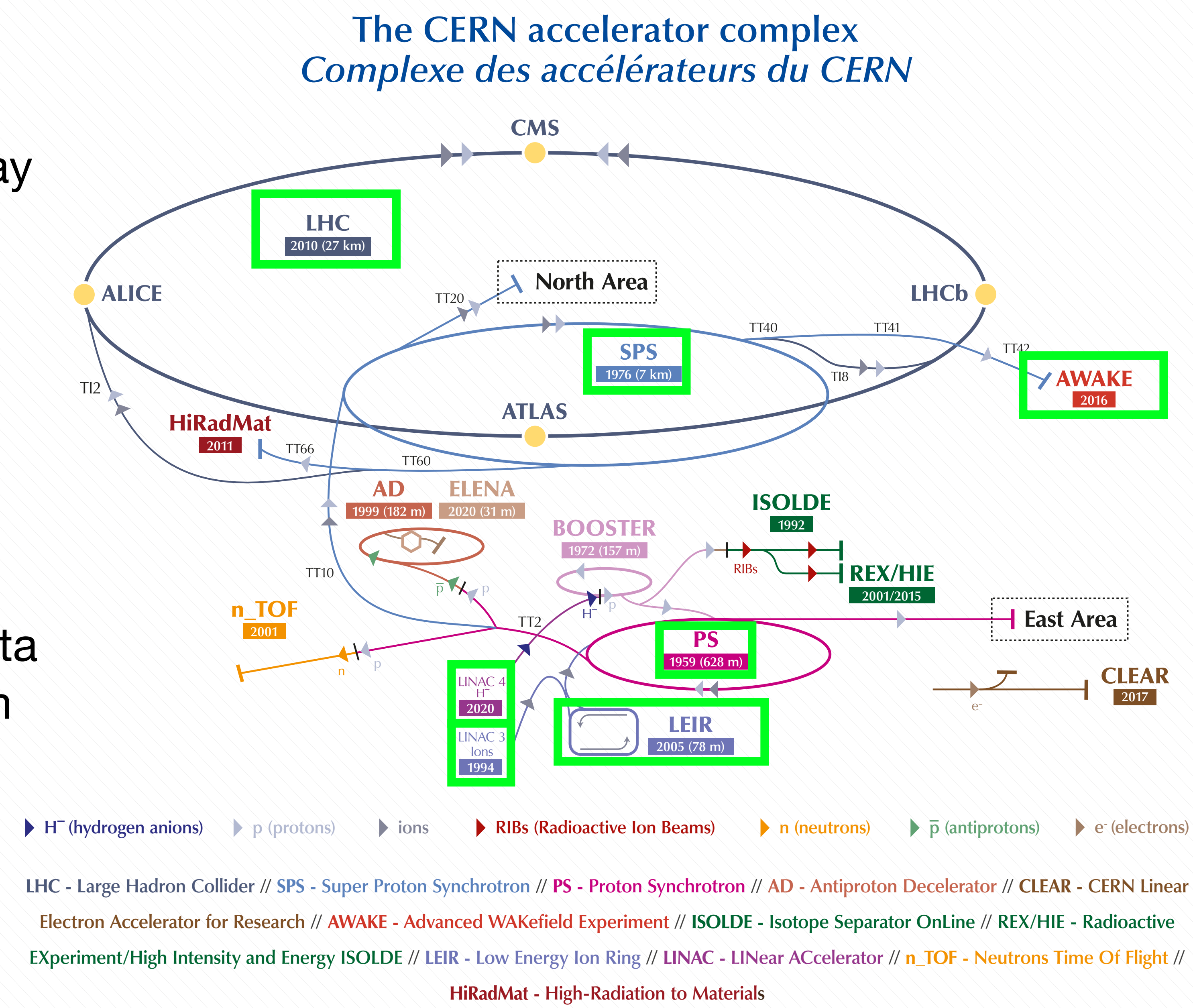


How CERN works

https://www.youtube.com/watch?v=pQhbhpU9Wrg

CERN accelerator complex

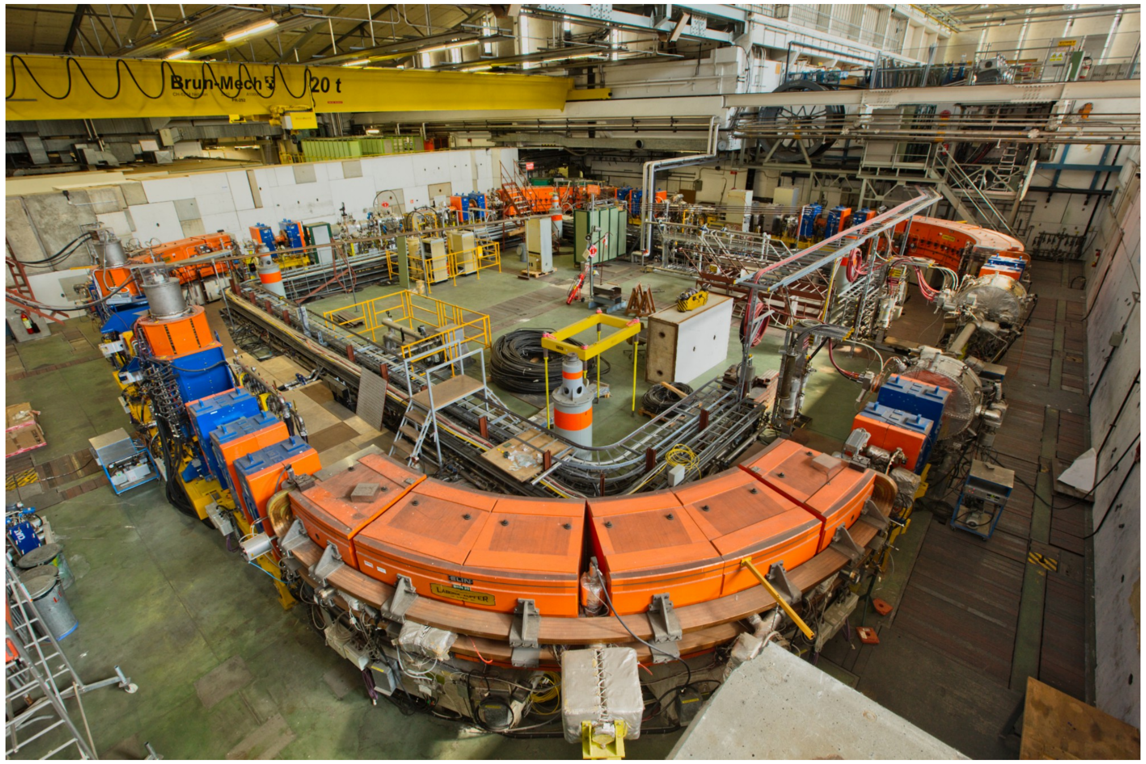
- Many challenges along the way
- Problem intrinsically hard to model:
    - Low energy as space charge in LINACs
    - Electron-cooling set-up
- Transmission-optimisation
- Alignment of electrostatic septa with many degrees of freedom
- …



How the story started: operating the Low Energy Ion Ring (LEIR)

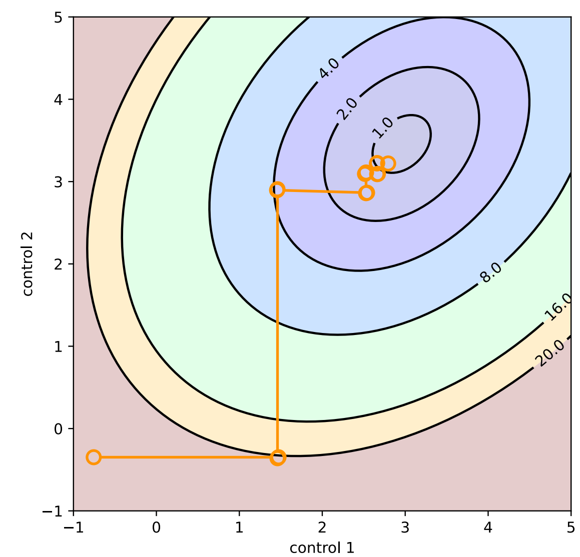
Supervision and operation:

- Complex system per design
- Many hours of manual maintenance/recovery of performance
- Introduction of automatic optimisation

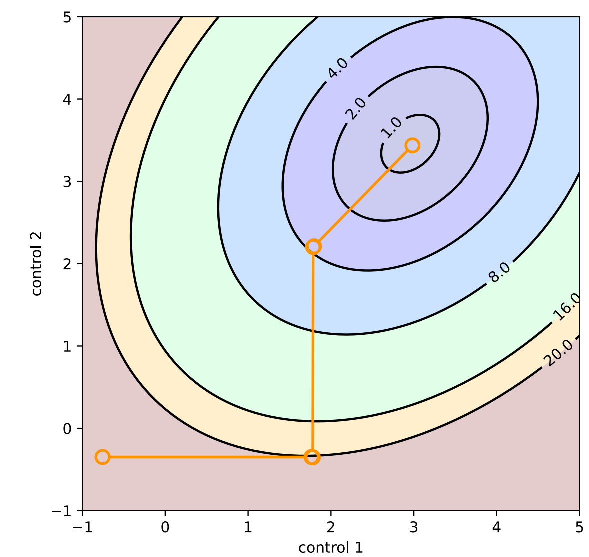


The raise of numerical optimisers

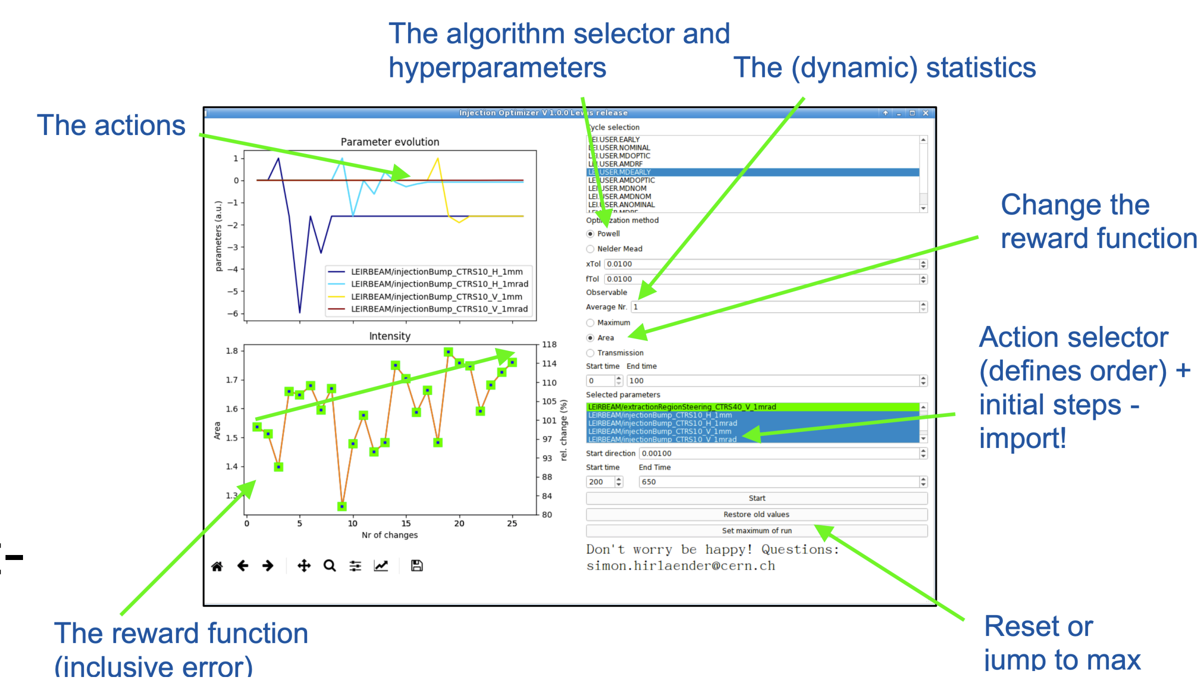
Manual



Optimiser



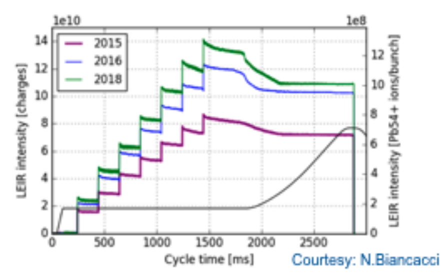
- Use of classical derivative free optimisers: Powell, Simplex, etc… (from ~1960)
- Simple UIs, scaleable, robust…
- <u>Enormous success</u>
- Reducing operations from hours manual steering to below one hours automatic set-up in below one hour



Powell 1964 - Optimisation

Achievements - LEIR

- 2018: record injected intensity into LEIR (and LHC)
- Fast recovery after LEIR machine stops and drifts
- Reproducible performance

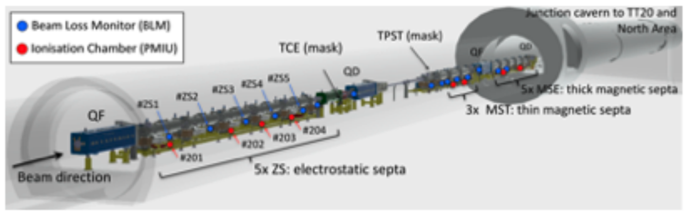


Courtesy: N.Biancacci

Result LHC 2018 for LEIR extracted intensity

75 ns	Mean $/10^{10}$c	Typical $/10^{10}$c	LIU$/10^{10}$c
LHC run	8.9	9.4	8.8

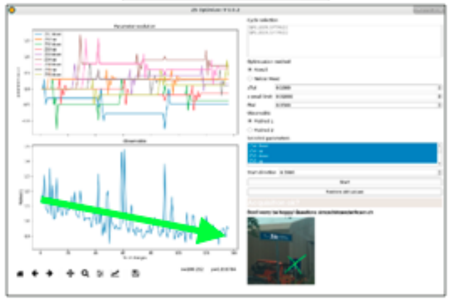
http://cds.cern.ch/record/2715365/



Normalized losses

Example: automatic alignment of electro-static septum for slow extraction at the SPS

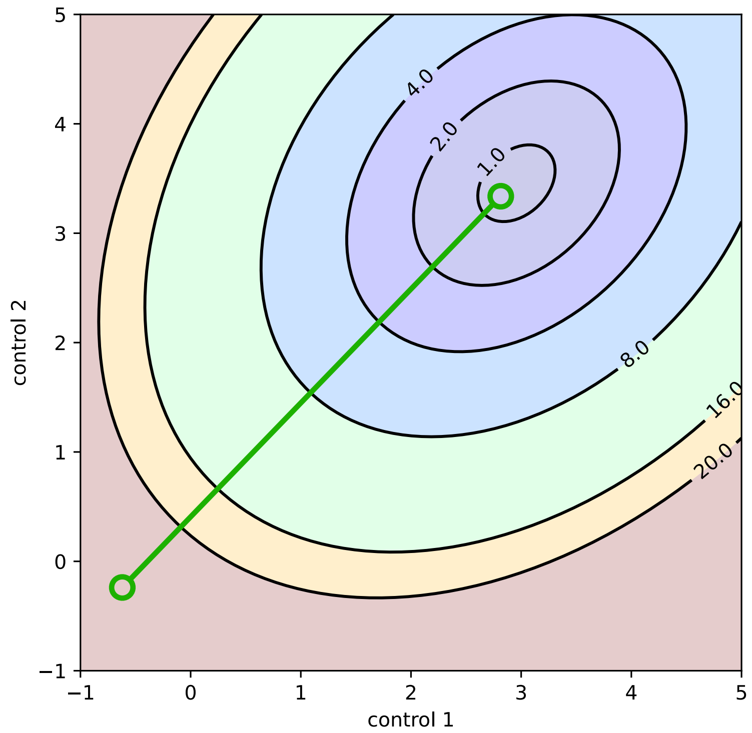
- 5 3.5 m long tanks with moveable anodes
  - 9 degrees of freedom to optimize; goal: minimize losses in extraction channel
  - Constrained to protect the hardware
- **Reduced alignment time** from ~ 8 h (quasi- manual scans) to ~ 45 minutes



https://doi.org/10.18429/JACoW-IPAC2019-THPRB080

Now optimisers in all flavours are standard tools

Beyond classical optimization: Reinforcement Learning

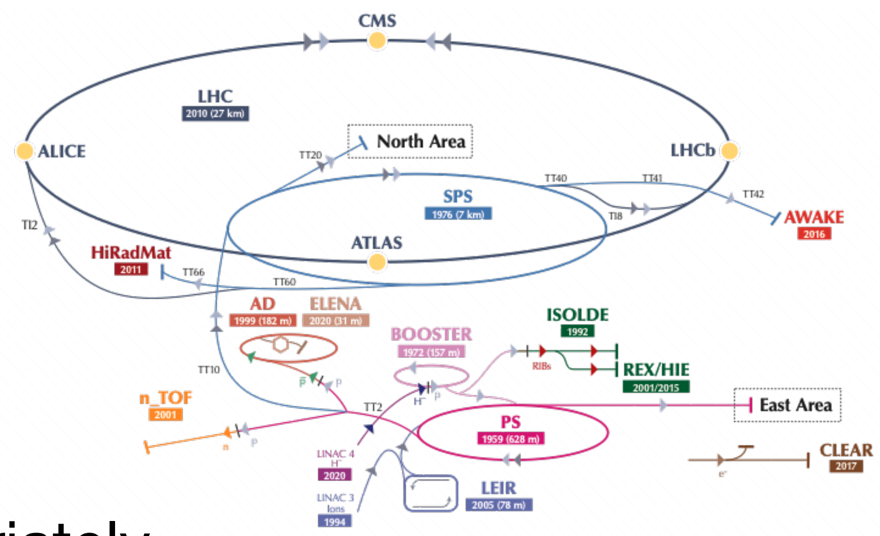


- Optimisation problems not solved from scratch each time from the beginning
- Existing data can be used
- Possible insights into the underlying physical problem
- Bigger class of problems can be addressed

https://indico.psi.ch/event/6698/contributions/16532/

Challenges of RL in accelerator control

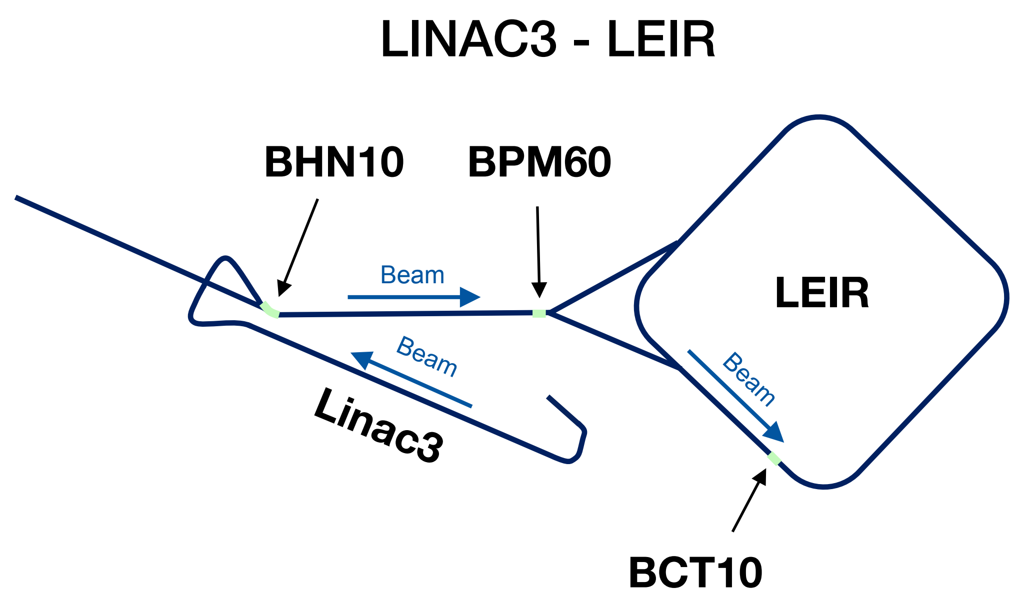
- Goal:
  - ➡ Quickly establish/recover performance
  - ➡ Maintain performance
- Challenges:
  - ➡ Not all processes can be modelled appropriately
  - ➡ Especially in the low energy regime lack of models
  - ➡ Accurate models are slow
- State representation sufficient for learning (beam diagnostics)?
  - ➡ Generally partially observable Markov decision processes (POMDPs)
- Sample efficiency - real world training feasible?
- Stability sufficient for real world training?
- Safety constrains?



Outline

- Motivation for RL and intro to RL
- What is CERN and why RL is interesting there
- **History of RL and examples**
- Resume and open questions

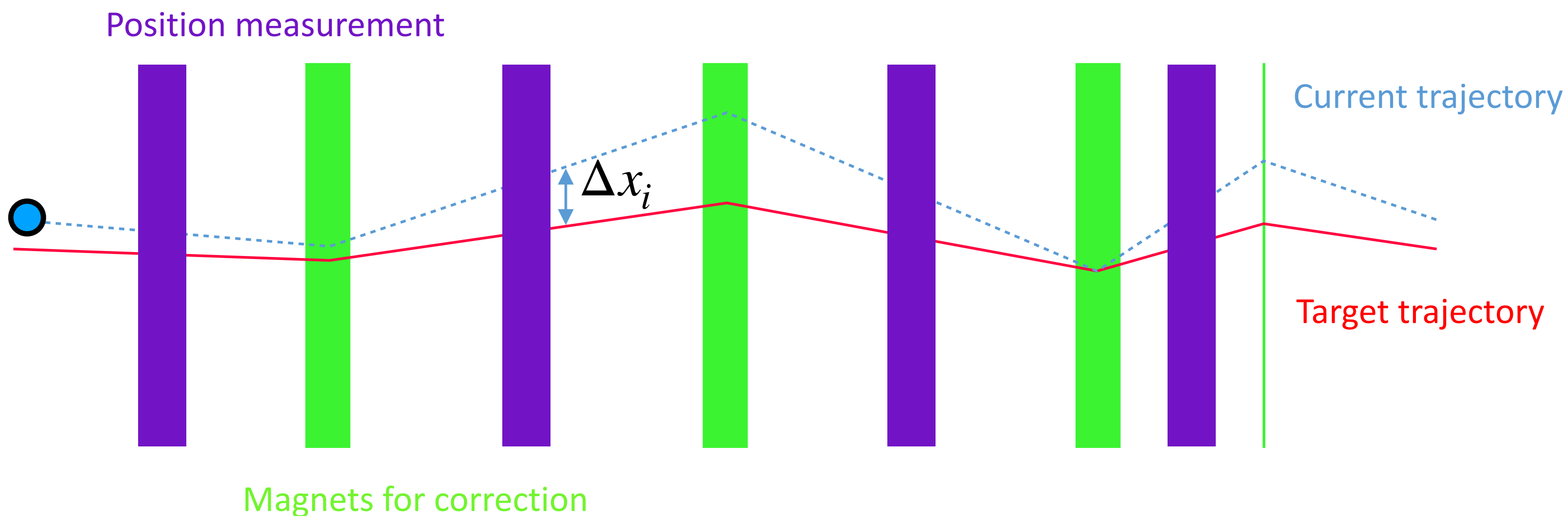
Starting with RL

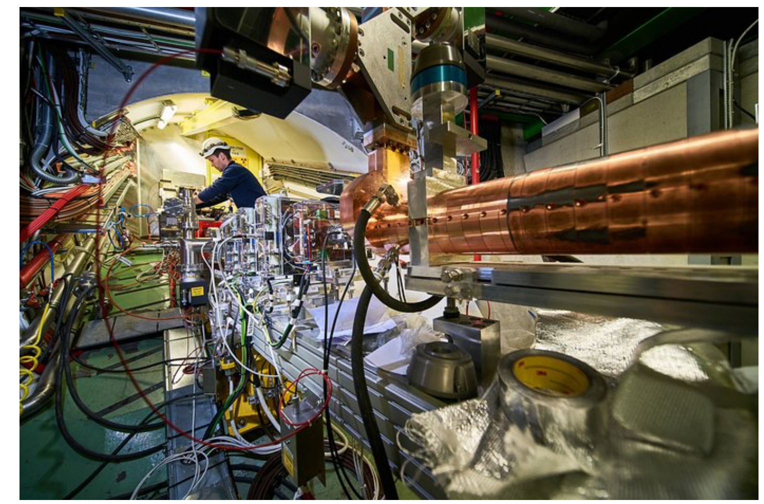


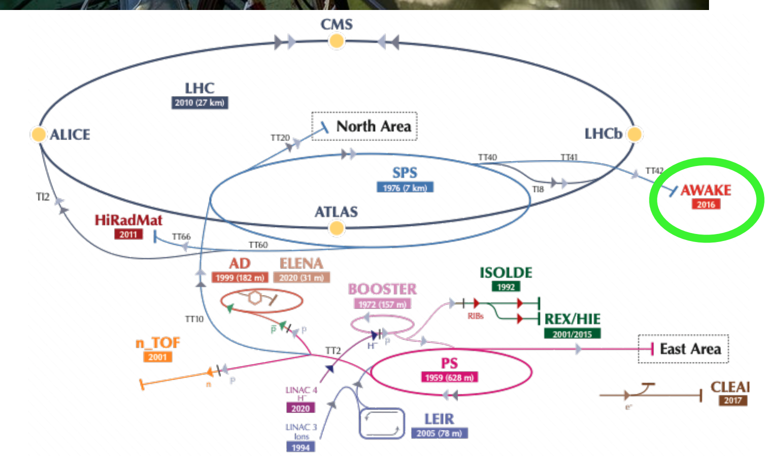
- 2018: Implementation of first deep reinforcement learning algorithm @ LEIR - proof of principle
- Challenges from infrastructural side
- Proof of principle experiments
- Starting benchmarking on AWAKE (Advanced Wake Field Experiment) trajectory steering

Benchmark: AWAKE trajectory steering

Accurate model







State = $\{\Delta x_1, \Delta x_2, \ldots \Delta x_{10}\}$

$\Delta x_i := x_{i_{\text{current}}} - x_{i_{\text{target}}}$

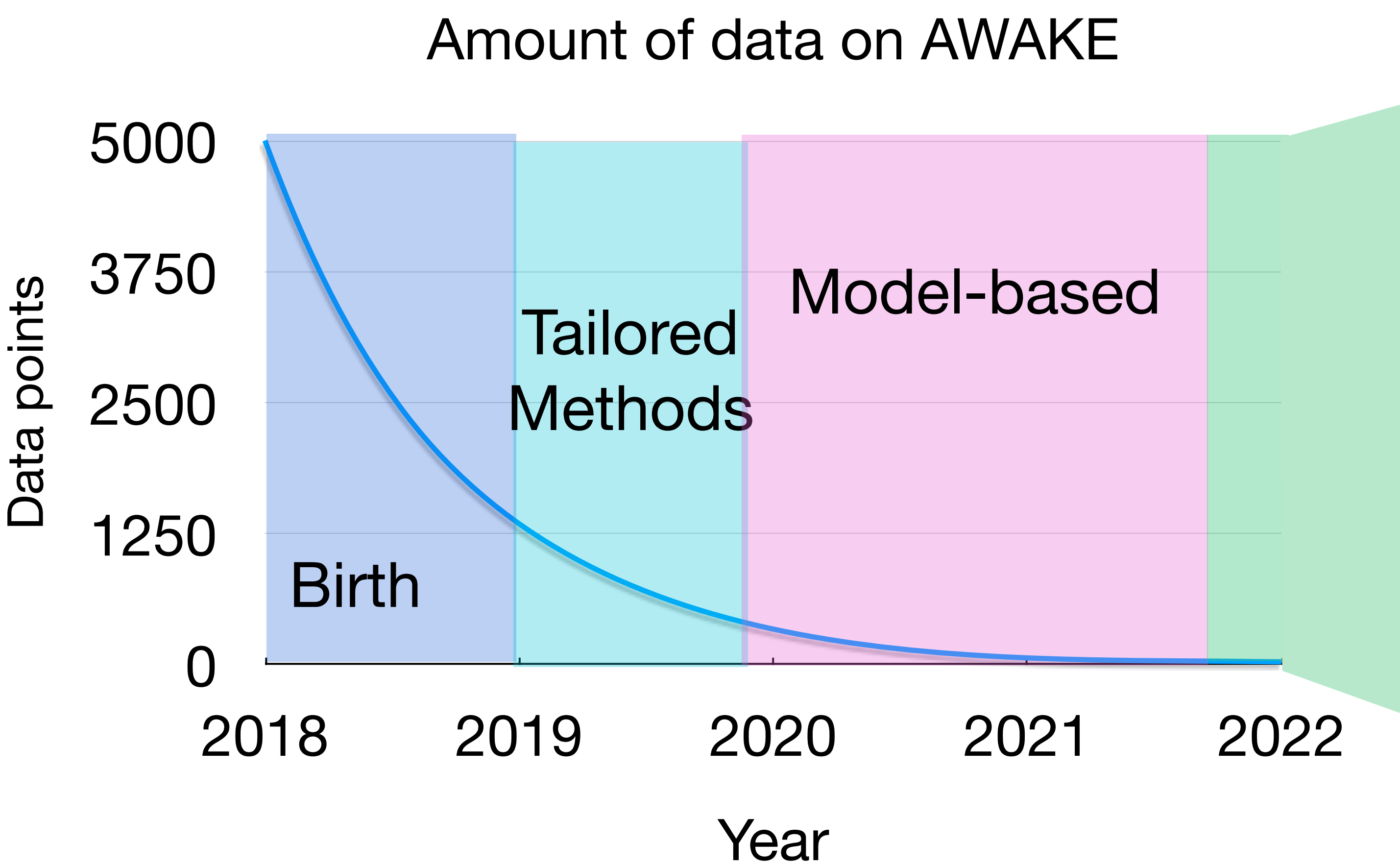
Actions = $\{k_0, k_1, k_2 \ldots, k_{10}\}$,

limited k_{max}

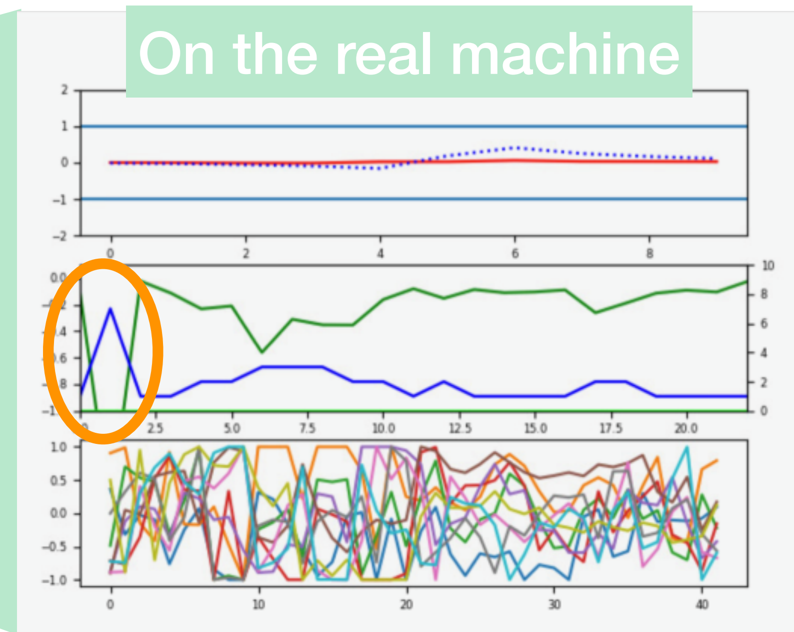
$$\boxed{\text{Reward} \propto -\sum_i^N \Delta x_i^2}$$

Target: trajectory steering - correct the trajectory in as little steps as possible.

Sample efficiency of RL on AWAKE



Shallow methods



Shown at IPAC 2023

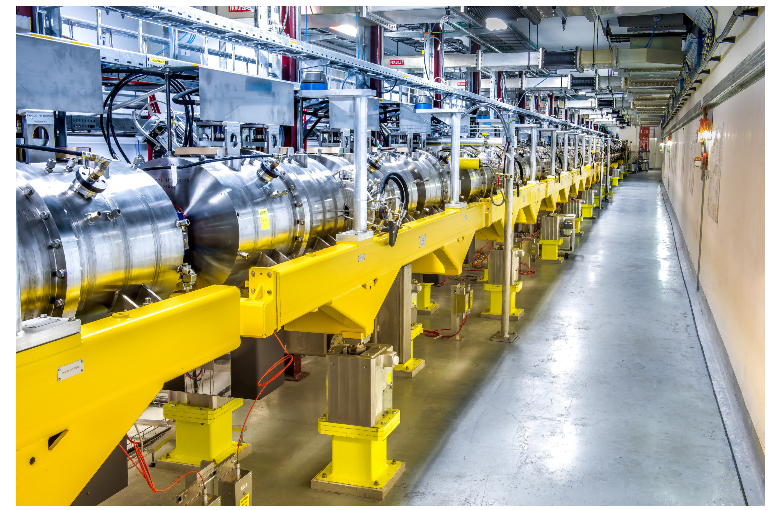
- Ultra fast reinforcement learning
  - ➡ Model-based using Gaussian processes
  - ➡ Only a few steps on the real machine from scratch
  - ➡ Overcomes limitations as non-stationarity and safety

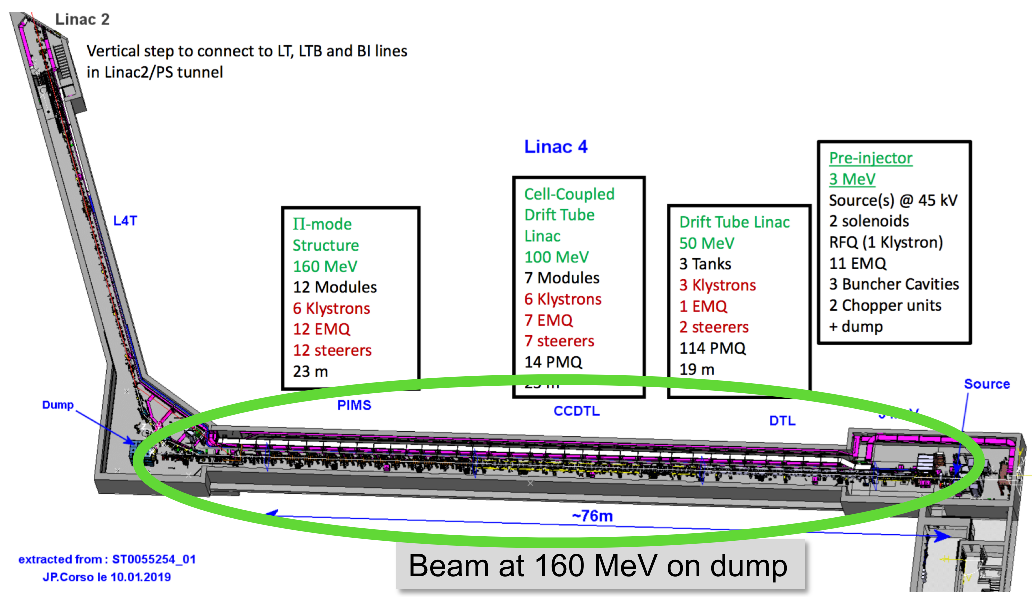
Off the shelf algorithms	Specific algorithms	Model-based Algorithms	Model-based Algorithms with convergence guarantees
PPO TRPO DDPG SAC TD3	NAF PER NAF	ME-TRPO Dyna-style	MBPO

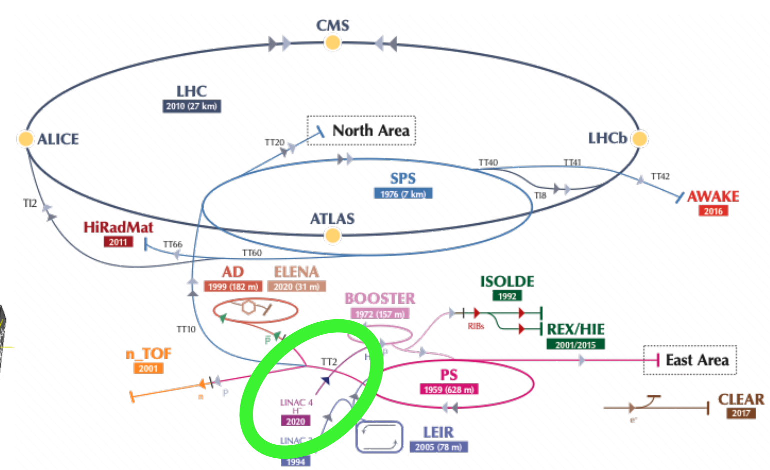
LINAC4 beam steering

LINAC4 (linear accelerator)

- 16 magnets
- H^{+} ion beam
- 76 m







Tailored Deep Q-learning ~100-250 steps

Model-based Dyna-Q ~100 steps

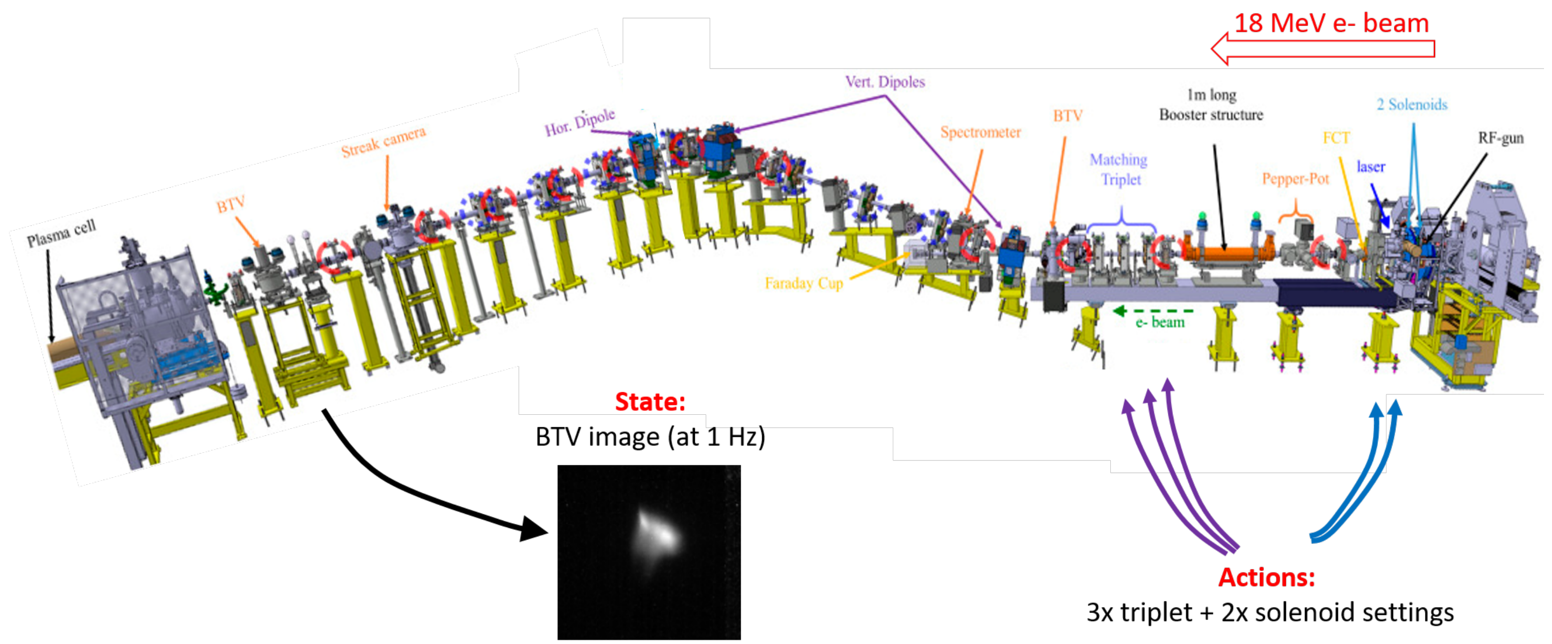
~ 100 iterations

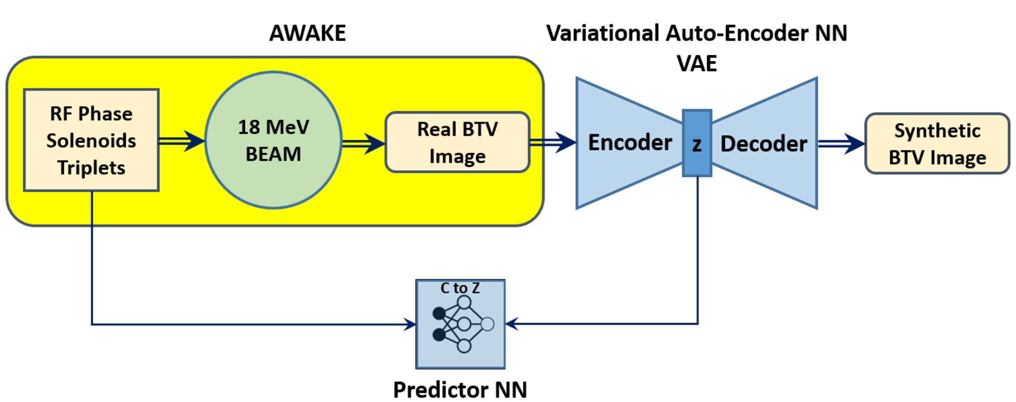
> 1e6 iterations

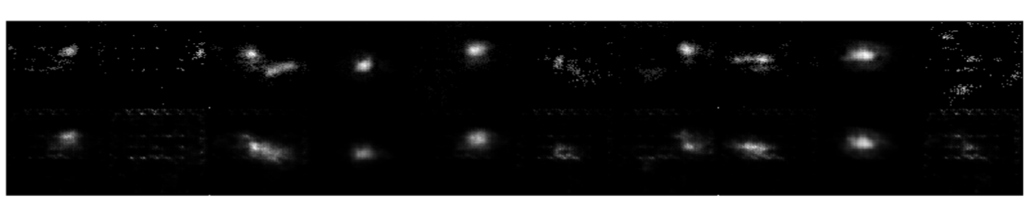
Model-based | Off-policy Q-learning | Actor-critic | On-policy Policy gradient | Evolutionary/ gradient free

https://journals.aps.org/prab/abstract/10.1103/PhysRevAccelBeams.23.124801

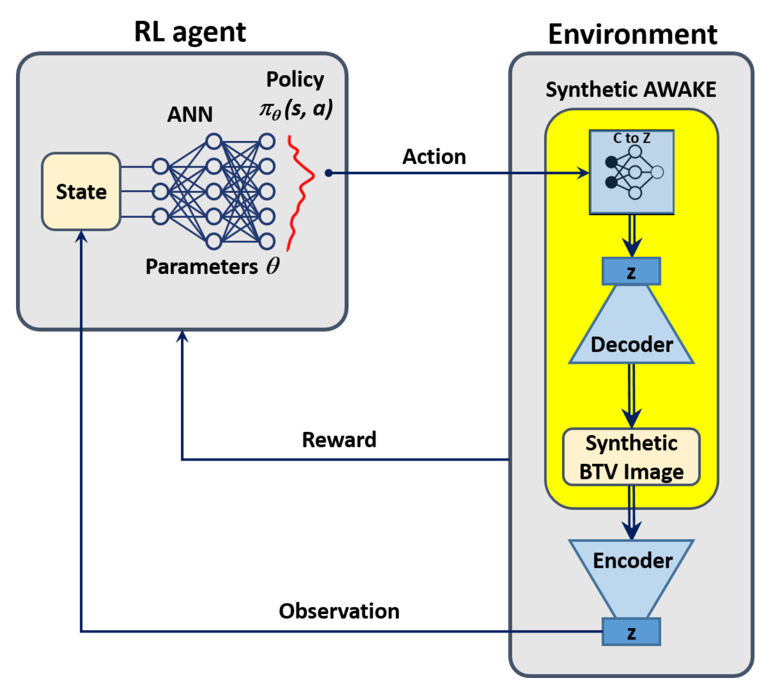
Deep fake AWAKE
Learning from (synthetic) images





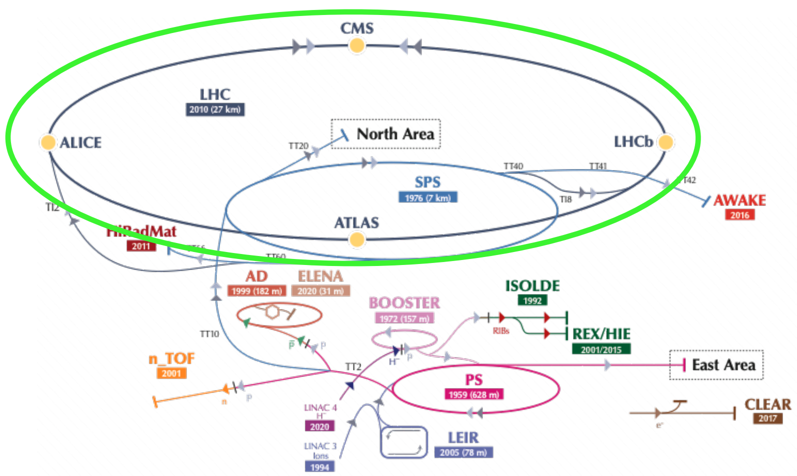


Synthetic images

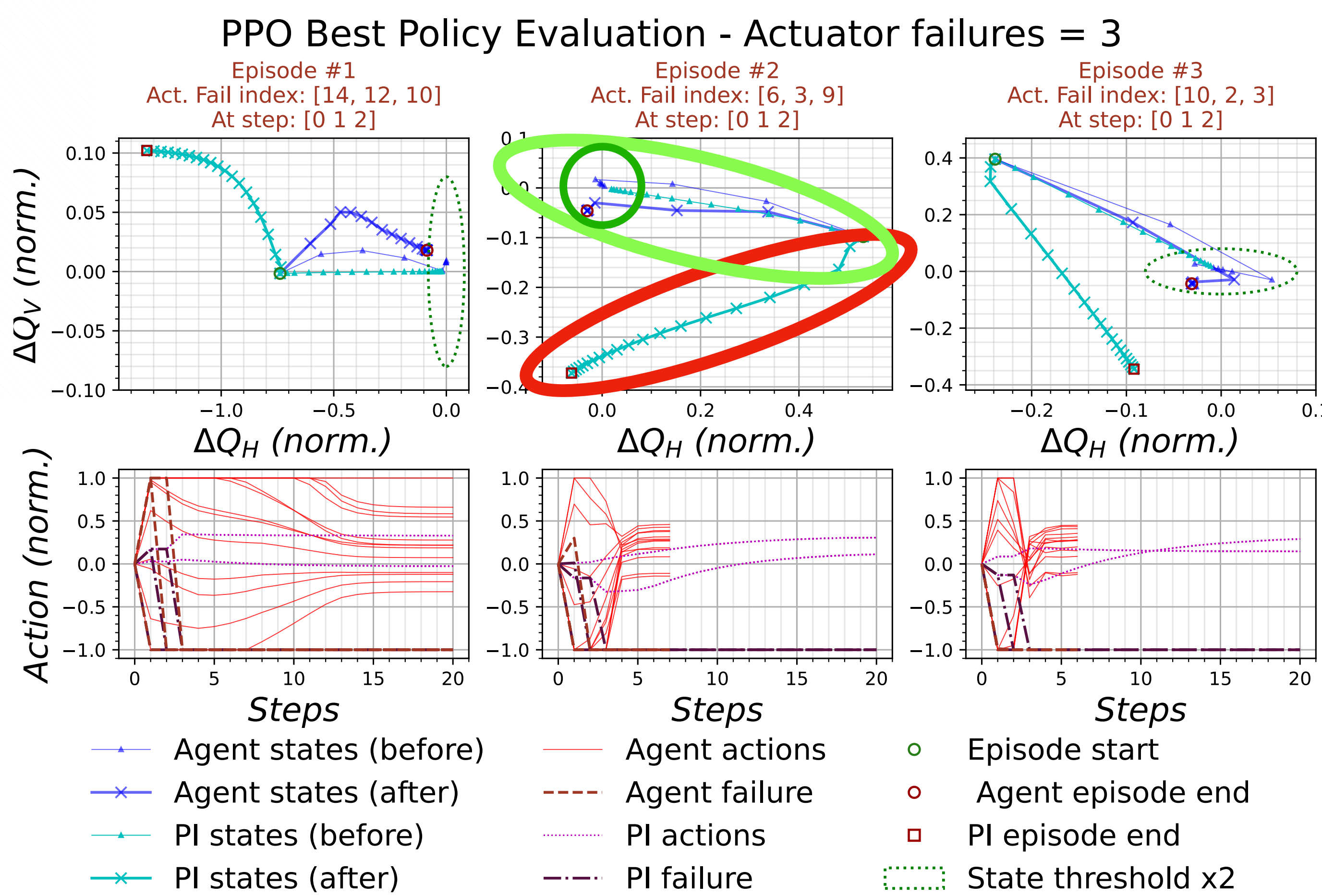


https://arxiv.org/abs/2209.03183

LHC Tune Feedback - beyond classical control
What can an RL agent do better?



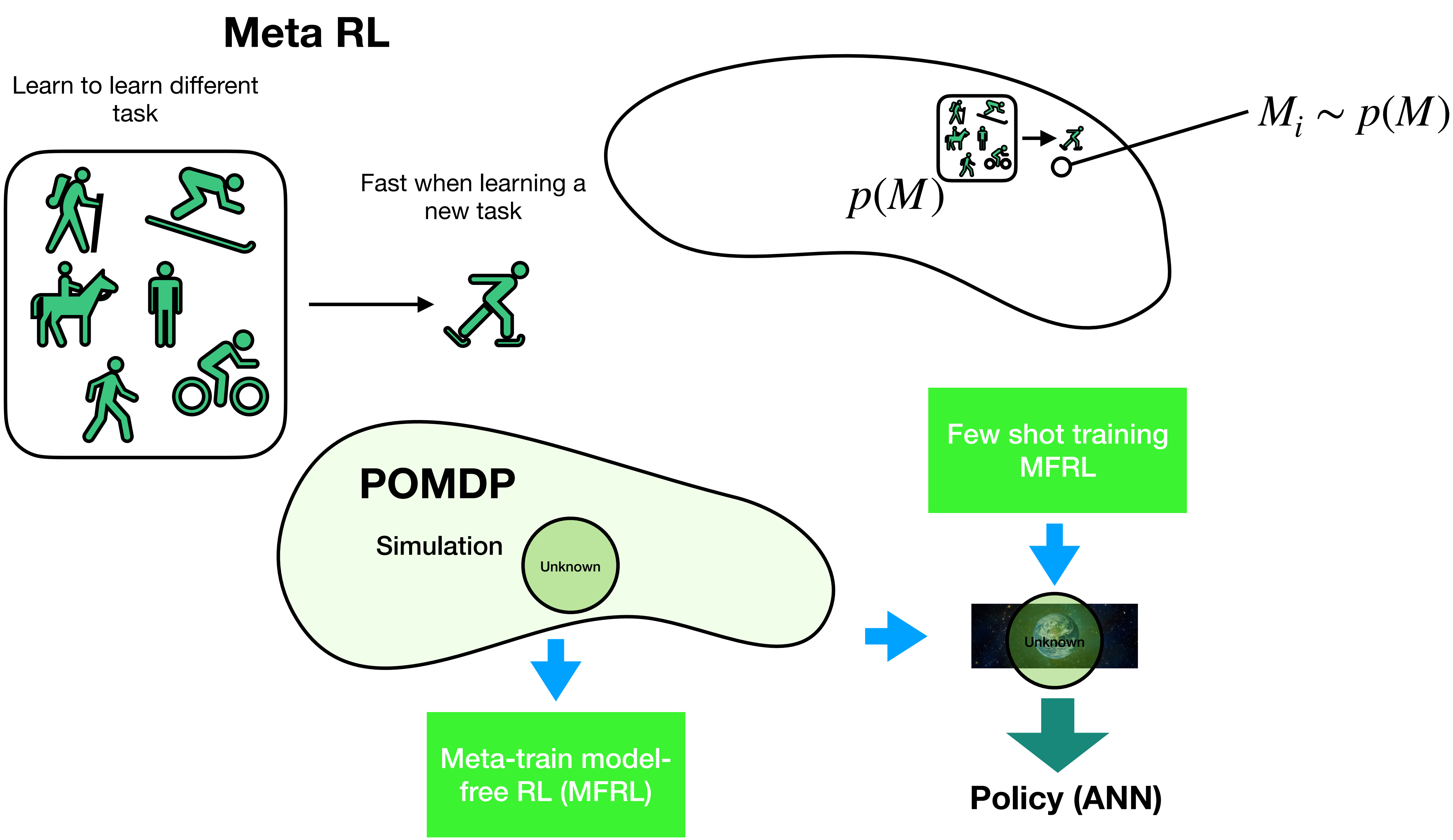
- Circular accelerator with Eigenfrequency Tune Q
- Currently: PI-controller
- 16 magnets
- Minimise ΔQ
- Simulation



https://www.frontiersin.org/articles/10.3389/fphy.2022.929064/

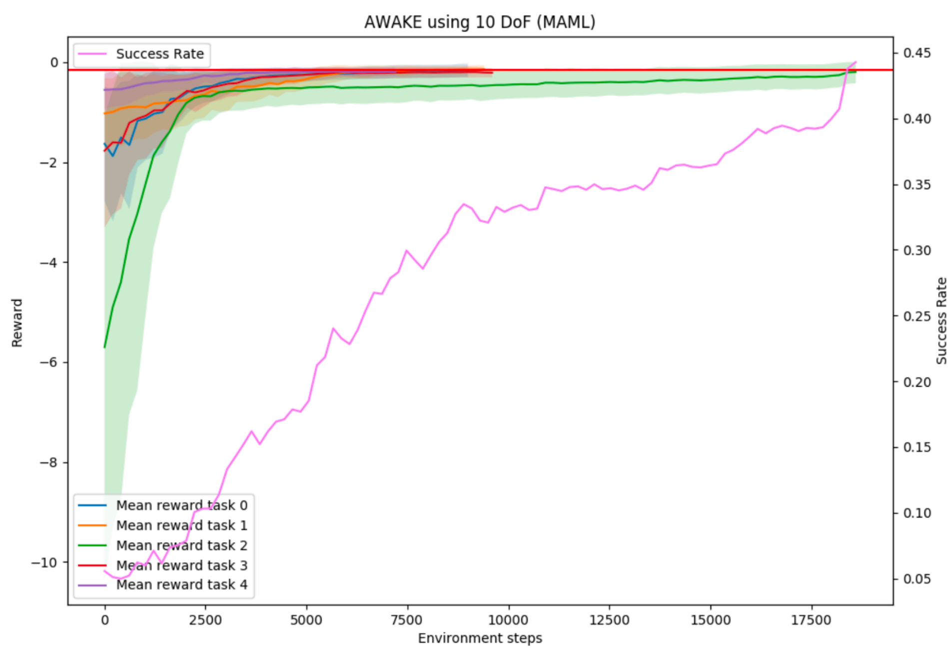
Beyond classical paradigms

- Learning to learn reinforcement learning

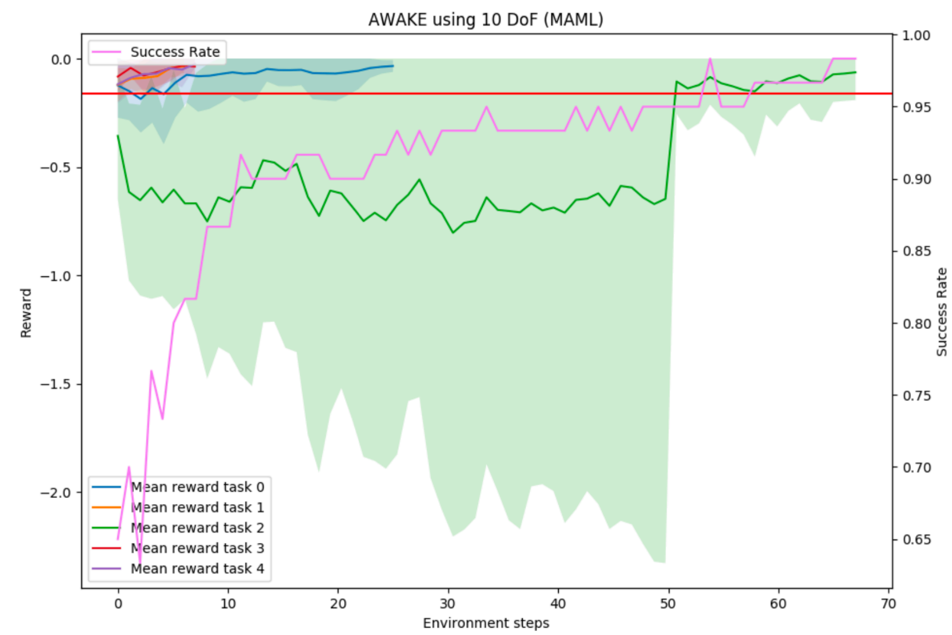


Meta Reinforcement Learning

- Learn on a distribution of tasks (high fidelity simulations) on AWAKE - 10 magnets, varying the quad-strengths
- Using a stable and monotonic algorithm
- Adapt quickly to actual setting - few shot adaption



Untrained ~ 18000 samples 40% success



Meta-trained ~80 samples 100% success ~ **few steps on the machine**

~ 100 iterations

> 1e6 iterations

Model-based

Off-policy
Q-learning

Actor-critic

On-policy
Policy gradient

Evolutionary/
gradient free

Demonstrated on the machine

Work with Lukas Lamminger Kain Verena

Outline

- Motivation for RL and intro to RL
- What is CERN and why RL is interesting there
- History of RL and examples
- **Conclusion and open questions**

Is RL the right tool?

- Optimisers:
  - ➡ Always re-explore - no memory → RL can
  - ➡ Cannot handle delayed consequences → RL can
- Accelerators seem to be generally a good environment for RL:
  - ➡ Generally known reward e.g. intensity (nevertheless might hard to design)
  - ➡ The state defined through beam diagnostics
  - ➡ The actions are mostly well designed
- Open issues:
  - ➡ What if no sufficient state available?
  - ➡ How to deal with non-stationarity?
  - ➡ How to improve the sample efficiency?
  - ➡ Stability - how to tune the algorithms?
  - ➡ What about safety?

What has changed?

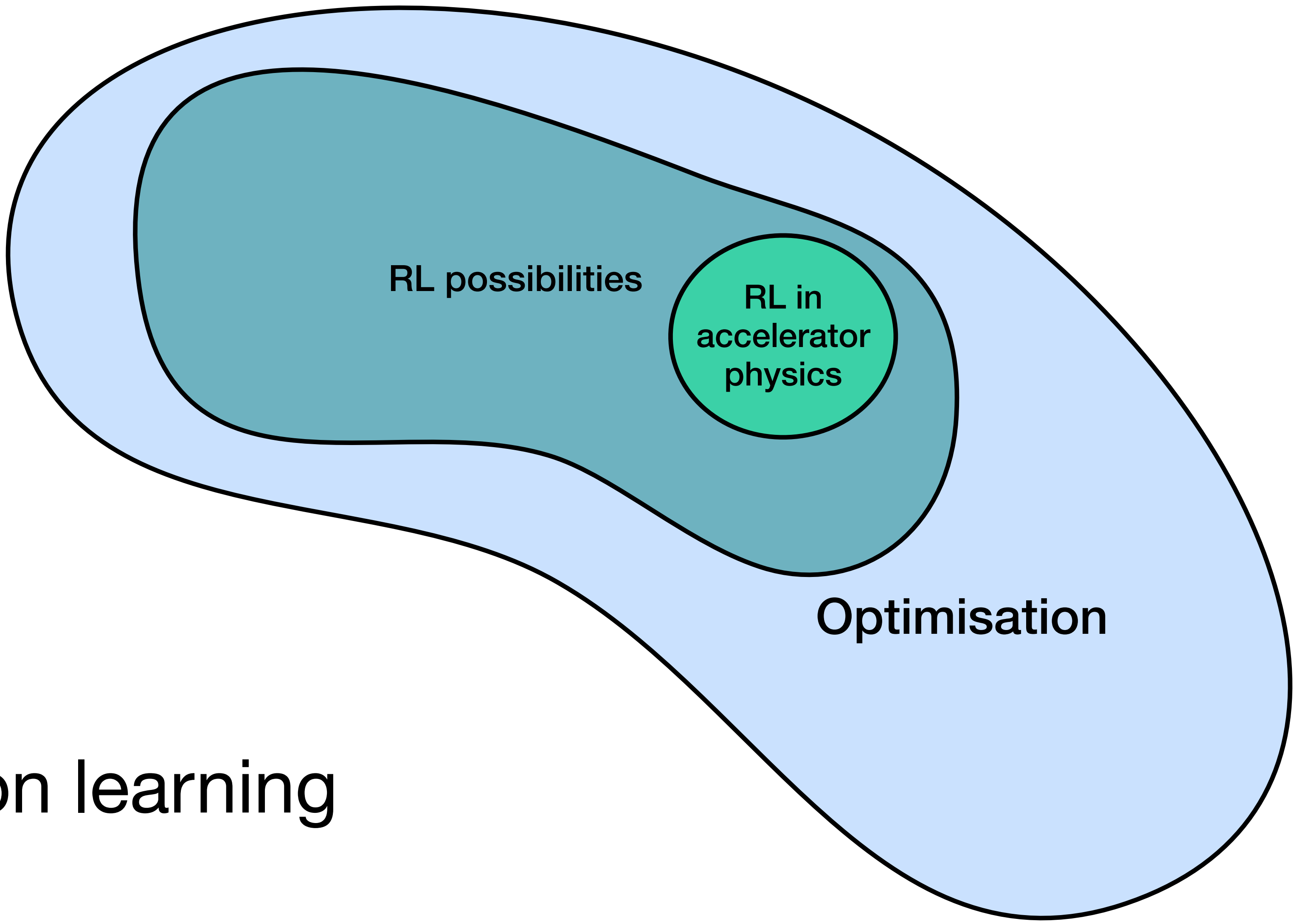
- Ecosystem and infrastructure has been established - modular systems - no general solutions
- We start to master many challenges:
    - ➡ Sample efficiency, safety, stability…
- We are not using the full potential of RL

We should use RL beyond optimization acceleration!

- (Model-based) Optimization replaced by RL
- Optimization is greedy!
- We don't leverage the full power of RL
- RL has another goal

Other avenues still to explore…

- Meta RL
- Multi task RL
- Contextual RL
- Multi-agent RL
- Hierarchical RL
- Distributional RL
- Inverse RL/Imitation learning
- …



Why is RL not applied more often?

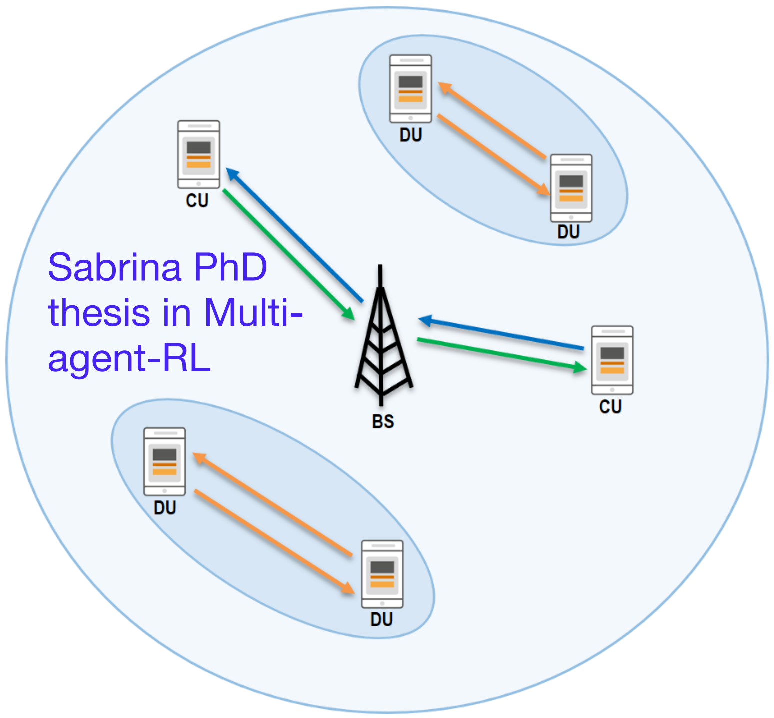
- General - not specific to accelerators
- RL is specific as many machine learning solutions
- Active paradigm:
  - ➡ Training and evaluations are challenging
  - ➡ Needs some experience
  - ➡ Rethinking of classic approaches as optimisation
- Still mainly a research topic than a standard approach
- What can we do?

More events like this!

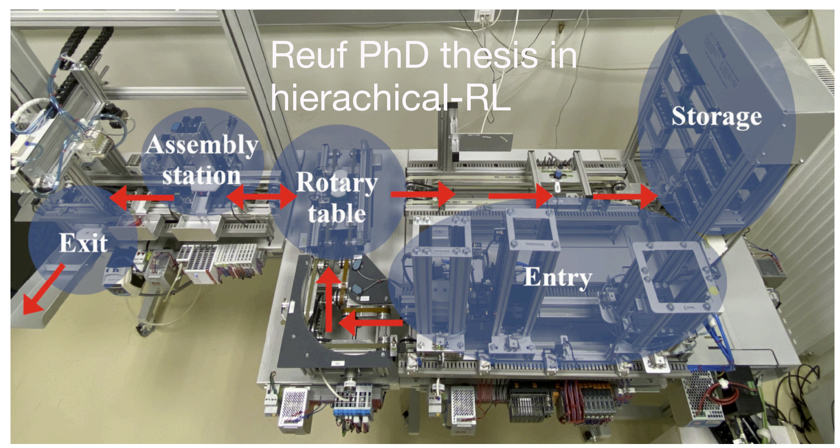


Build a stronger community
Collaborate more

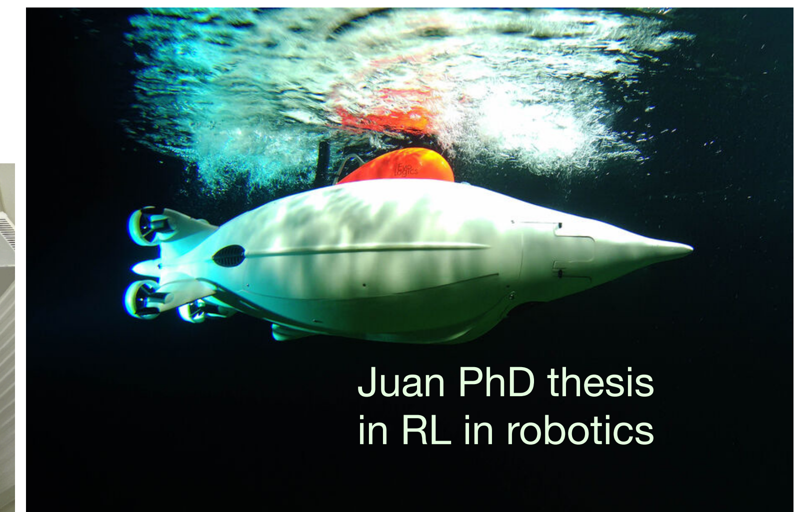
What “my” RL students do



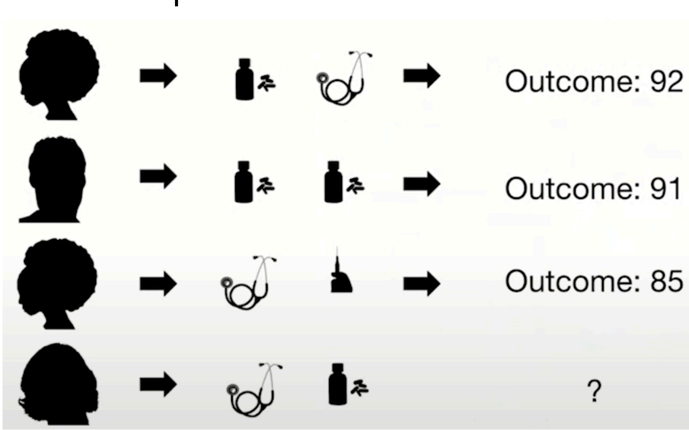
Sabrina PhD thesis in Multi-agent-RL

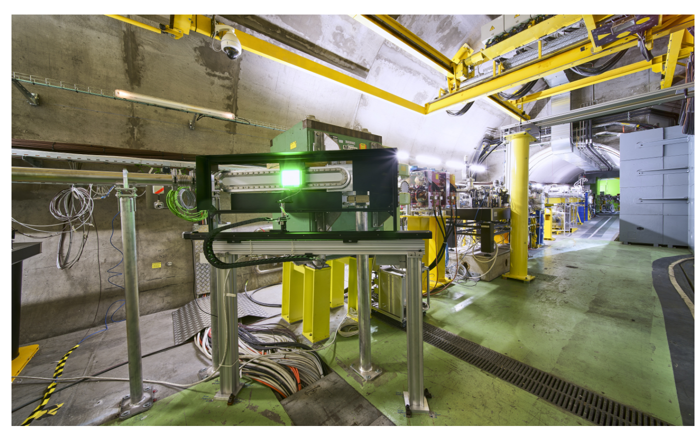


Reuf PhD thesis in hierachical-RL


Juan PhD thesis in RL in robotics

Sascha master thesis in interpretable RL in medicine

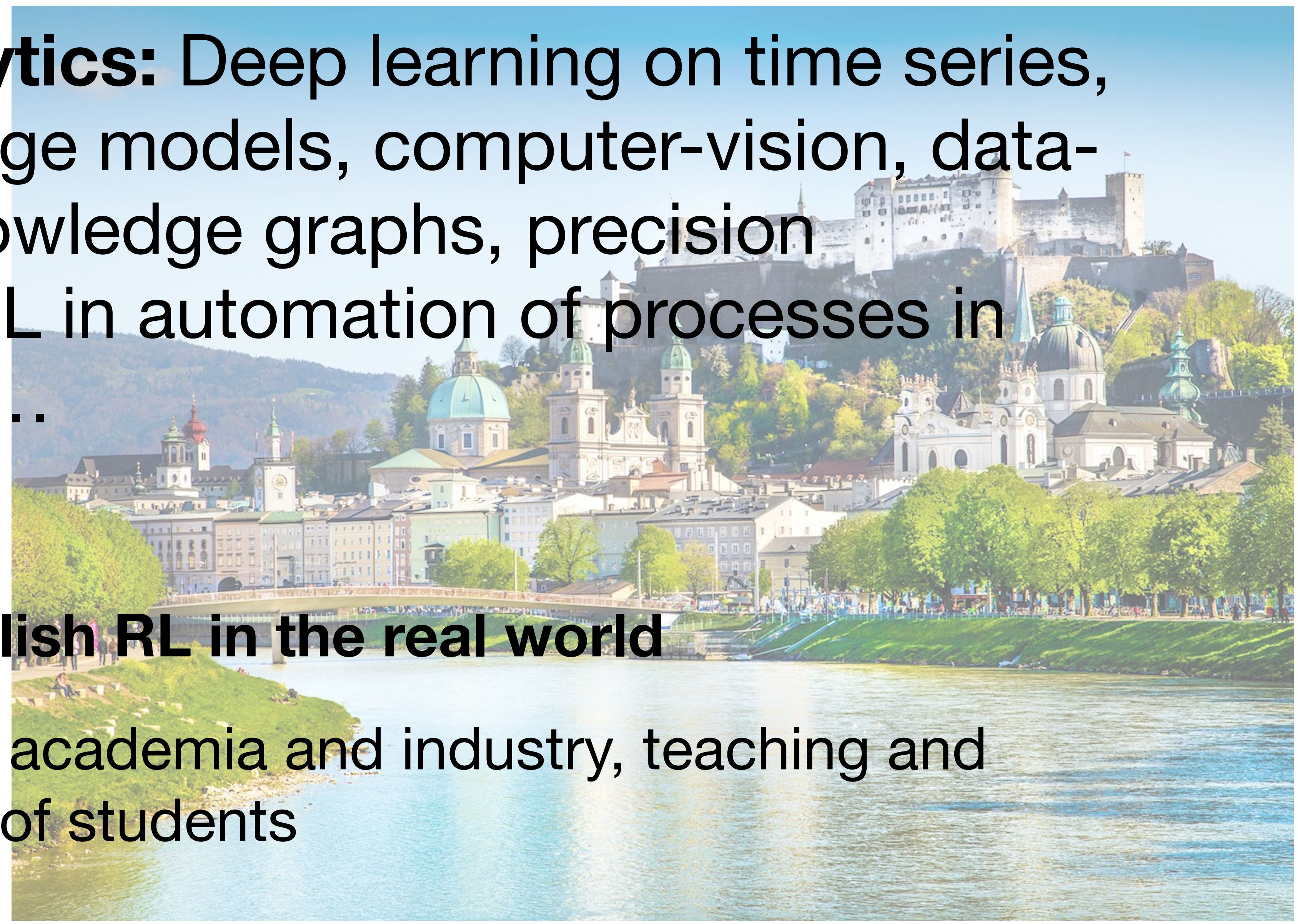



Lukas master thesis in Meta RL

Myself Industry reduce power consumption of companies

Thanks for your attention

My team: Smart Analytics und Reinforcement Learning - IDA Lab



- **Smart analytics:** Deep learning on time series, large language models, computer-vision, data-science, knowledge graphs, precision medicine, ML in automation of processes in companies,…
- **RL**:
  - ➡ Goal: **Establish RL in the real world**
  - ➡ Research in academia and industry, teaching and supervision of students